U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

MSc

2.º CICLO

FCUP
2016

U.PORTO

Computational studies addressed to Histidine decarboxylase

Henrique Silva Fernandes

FC

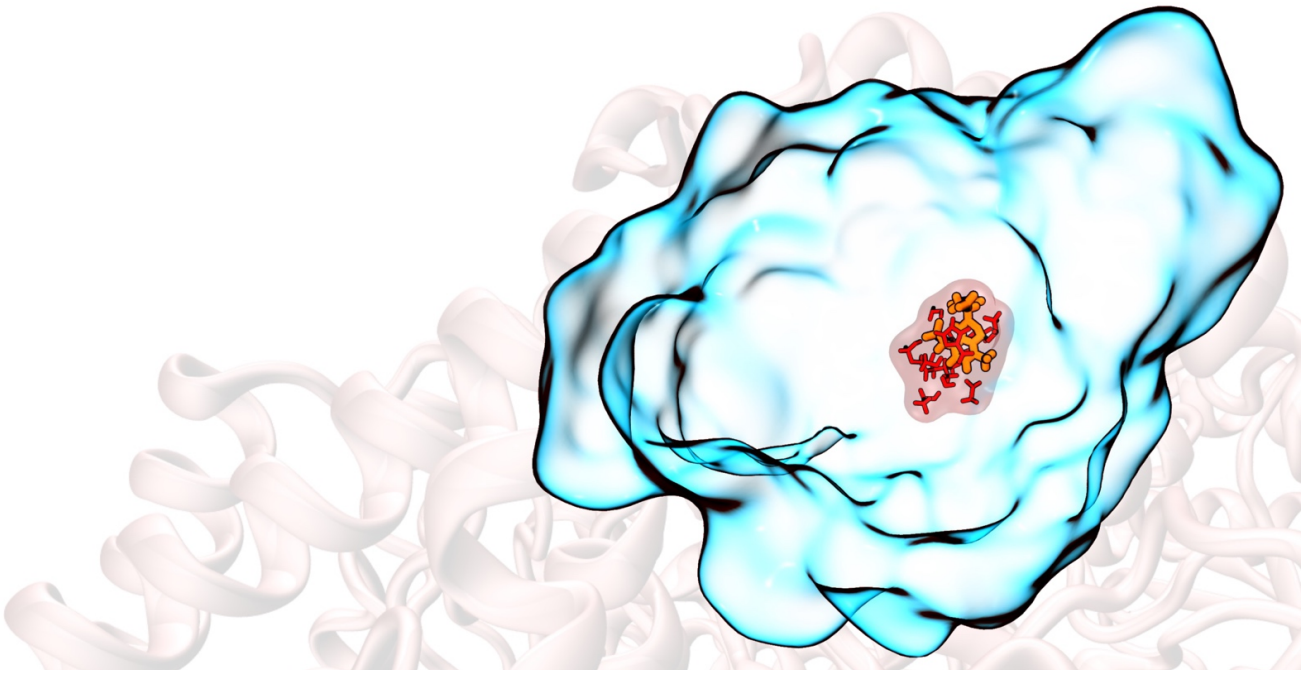
U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

# Computational studies addressed to Histidine decarboxylase

Henrique Silva Fernandes

Dissertação de Mestrado apresentada à Faculdade de Ciências da Universidade do Porto em Bioquímica

2016



# Computational studies addressed to Histidine decarboxylase

Henrique Silva Fernandes
Mestrado em Bioquímica
Departamento de Química e Bioquímica
2015/2016

**Supervisor**
Nuno Manuel Ferreira de Sousa de Azevedo Cerqueira, Assistant Researcher,
REQUIMTE, Faculdade de Ciências da Universidade do Porto

U.PORTO

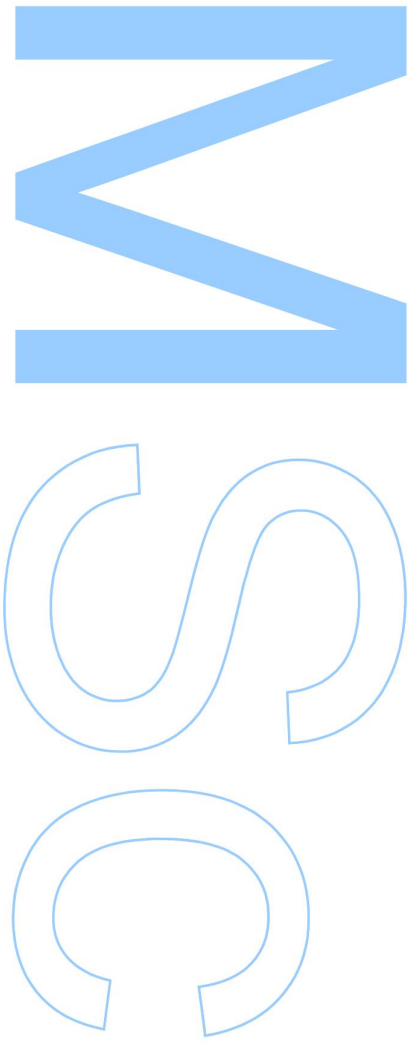
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



Esta dissertação foi realizada no UCIBIO-REQUIMTE, no Grupo de Química e Bioquímica Teórica e Computacional do Departamento de Química e Bioquímica da Faculdade de Ciências da Universidade do Porto, sob a orientação do investigador Nuno Manuel Ferreira de Sousa de Azevedo Cerqueira.

Este projeto teve ainda o apoio financeiro da Fundação para a Ciência e Tecnologia (FCT), através do projecto IF/01310/2013, e dos fundos nacionais e co-financiamento da FEDER, sob o acordo de parceria PT2020.

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

UCIBIO
@ REQUIMTE

requimte
rede de química e tecnologia

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

IF INVESTIGADOR FCT
Investigador FCT 2013
Candidaturas abertas

This page was intentionally left blank.

# Agradecimentos

Esta tese assinala o término de mais uma importante etapa da minha vida académica, e é com muito gosto que agradeço a todos os que contribuíram para que esta fosse possível.

Em primeiro lugar agradeço imenso ao Professor Doutor Nuno Cerqueira, pela excelente oportunidade que me concedeu para trabalhar consigo. Foi um ano incrível, muito produtivo (na minha opinião) e onde aprendi imenso. Não posso deixar de agradecer toda a sua simpatia, empenho, apoio, dedicação, ensinamentos e enorme disponibilidade que demonstrou ao longo de todo o projeto. Saliento também o contínuo espicaçar para fazer trabalhos diferentes e abordar os problemas considerando diferentes perspetivas.

À Professora Doutora Maria João Ramos, agradeço a possibilidade de integrar este fantástico grupo de investigação, os ensinamentos e apoio manifestados ao longo da minha dissertação.

Ao Professor Doutor Pedro Alexandrino Fernandes, que na qualidade de diretor de mestrado, manifestou-se sempre disponível para esclarecer as dúvidas que iam surgindo.

A todo os elementos do grupo da Química Teórica e Bioquímica Computacional, pela forma acolhedora e amizade com que me receberam, e que também contribuíram para o sucesso deste trabalho. Em especial, agradeço à Carla, minha “colega de carteira”, que me ajudou muito na adaptação durante os primeiros tempos e que continua a propiciar uma aprendizagem contínua e recíproca. Um agradecimento muito especial também a todos os maravilhosos colegas da sala 3.26, Elizabeth, Inês, Carla, Kris, Marco (o excelente Pasteleiro) e Rui pelos bons momentos de descontração e conversas acerca da “atualidade”.

Ao Óscar, por todo o apoio informático prestado quando as coisas funcionavam menos bem.

Ao Rui Neves, João, Rita, Cátia e Fabiola pelas interessantes discussões durante o almoço, em particular ao Rui que contribuí também com os seus ensinamentos para este projeto.

Aos Maiores, um muito obrigado ao Nuno, Chaves e Ricardo (desaparecido) por toda a amizade, momentos de trabalho e descontração. Tem sido excelente poder partilhar todos estes anos convosco.

À minha irmã, pelo apoio, amizade e admiração.

Aos meus avós (Avó Ilda, Avó Idalina e Avô Afonso), que sempre quiserem o melhor para mim e por isso sempre me incentivaram e apoiaram durante estes anos.

Um indiscritível agradecimento aos meus pais, por todo o esforço que fizeram e fazem para que eu concretize os meus sonhos. Obrigado por todo o apoio, incentivo e dedicação que sempre demonstraram. A vocês dedico esta tese.

Por fim, e de forma muito especial, agradeço-te a ti, Jéssica, por toda a compreensão, paciência e encorajamento manifestados. Sempre estiveste lá ao meu lado, sempre acreditaste em mim, sempre me ajudaste nos momentos difíceis e sempre soubeste. Por isso, também te dedico a ti, esta tese.

# Abstract

Histamine is a very important biological amine that plays a key role in several biological events such as immune response, gastric system modulation and also as a neurotransmitter in the central nervous system. However, there are some disorders associated with an overproduction of histamine, and, consequently, several inhibitors for histamine action have been studied and developed in order to treat them. Among all histamine-related diseases, we can highlight allergies, atopic dermatitis, and cancer. Since the *in vivo* production of histamine is catalyzed by Histidine decarboxylase (HDC), in this work, we studied the catalytic mechanism of this enzyme using computational means.

HDC is an enzyme that requires pyridoxal-5´-phosphate (PLP) as a cofactor. This enzyme belongs to the group II of PLP-dependent decarboxylases together with *L*-DOPA and glutamate decarboxylases, and catalyzes the *L*-histidine decarboxylation from which results histamine and carbon dioxide.

In this work, we studied the catalytic mechanism of HDC by computational approaches using the recent X-ray structure of HDC (PDB code: 4E1O) and an ONIOM QM/MM methodology. The QM part was calculated using DFT method (B3LYP with the 6-31G(d) basis set for geometry optimizations and the M06 functional with the 6-311++G(3df,2pd) for single point energy calculations), whereas the MM part was considered under MM approach using the GAFF and ff99SB force fields amongst other parameters determined by us.

The results showed that experimental data does not always give accurate and full information about the structure of biomolecular complexes. Therefore, computational methods are important techniques to test and complement these results. In the particular case of this work, an inhibitor was co-crystallized with the enzyme, but, apart from its similarity with the natural substrate, several issues were founded in the structure from PDB file that prevents the reaction to occur with acceptable energetics. During this work, we proved that such orientation of the external aldimine inside the active site is not suitable for catalysis under physiological conditions and that a different conformation needs to be taken into account. All the obtained results are consistent with the available experimental kinetic results. Then, we confirmed a two-step type of mechanism for the reaction catalyzed by the HDC. The first step involves a decarboxylation of the external aldimine (PLP cofactor bonded to the substrate), and it is followed by the formation of a stable carbanion (quinonoid intermediate). This is the rate-limit step, since the activation barrier is 12.70 kcal/mol, and it is also an endergonic step ($\Delta G_r$ = 10.52 kcal/mol). In the

second step, the quinonoid intermediated is protonated by Tyr-334B from which results histamine. This work proposed for the first time which amino acid residue from the active site was the one involved in this protonation stage. This step is faster than the first one ($E_a$= 2.87 kcal/mol) and, in contrast, is a high exothermic step ($E_r$ = -26.02 kcal/mol).

This work allowed a fully atomistic description of the catalytic mechanism of HDC, elucidating about the important role played by several amino acids from the active site, namely, Asp-273A, Lys-305A, Tyr-334B, His-194A, Ser-304A, Thr-248A, Asn-302A, Tyr-80A, Ser-354B. This data can now be useful for the development of new drugs to treat histamine-related diseases. Additionally, the importance of the computational means was consolidated once again, becoming a stronger approach to test and validate experimental data.

**Keywords**

Histidine decarboxylase, pyridoxal-5'-phospaphate (PLP), *L*-histidine, histamine, catalytic mechanism, quantum mechanics, hybrid QM/MM methodologies, ONIOM, enzymatic catalysis, density functional theory (DFT)

# Resumo

A histamina é uma amina que desempenha um papel importante em diversos fenómenos fisiológicos tais como a resposta imunitária, a modulação gástrica, atuando também como neurotransmissor no sistema nervoso central. Porém, existem várias doenças associadas a uma produção excessiva de histamina e, por isso, são vários os inibidores que têm sido estudados para atenuar os efeitos provocados por estas concentrações anormais de histamina nas células. De entre todas as doenças associadas a esta disfunção fisiológica, podemos destacar as alergias, a dermatite atópica e o cancro, como as mais preocupantes. Como a produção de histamina é catalisada pela Histidina descarboxilase (HDC), este trabalho foi focado no estudo do mecanismo catalítico desta enzima,.

A HDC é uma enzima que usa o piridoxal-5'-fosfato (PLP), a forma ativa da vitamine $B_6$, como cofator. Esta enzima pertence ao grupo II das descarboxilases que usam o PLP como cofator, juntamente com a *L*-DOPA e glutamato descarboxilases. A HDC é uma enzima muito específica que catalisa a descarboxilação da *L*-histidina, originando histamina e dióxido de carbono.

Neste projeto, estudámos o mecanismo catalítico da HDC usando como base a única estrutura da HDC disponível experimentalmente (código do PDB: 4E1O). Para este efeito foram utilizados métodos computacionais nomeadamente os métodos híbridos de QM/MM usando um esquema subtrativo (ONIOM). A parte QM foi calculada usando a teoria do funcional de densidade, o funcional B3LYP e a base de funções 6-31G(d), enquanto que a parte MM foi calculada por mecânica molecular com recurso aos campos de forças GAFF e ff99SB, mas também aos parâmetros que foram determinados por nós neste trabalho.

Os resultados mostraram que os dados experimentais nem sempre são perfeitos e podem dar informações incompletas acerca da estrutura de complexos biomoleculares. Por isso, as técnicas computacionais são uma excelente ferramenta para testar e complementar os resultados experimentais. No caso particular deste estudo, um inibidor foi co-cristalizado com a enzima, mas, apesar da sua similaridade com o substrato natural, os resultados computacionais mostram que a conformação adoptada por este e subsequente modificação pelo substrato natural impossibilitavam a reação em meio fisiológico. Consequentemente, durante este trabalho, esta estrutura foi sequencialmente modelada de forma a obter um modelo que fosse consistente com os dados cinéticos obtidos por técnicas experimentais. Após as devidas correções ao modelo, foi possível confirmar um mecanismo catalítico de dois passos para a reação

catalisada pela HDC. O primeiro passo envolve a descarboxilação da aldimina externa (o cofator de PLP ligado ao substrato), que origina um carbanião altamente estabilizado pelo cofator (intermediário quinonóide). Este é o passo limitante da reação, uma vez que a barreira de ativação é de 12,70 kcal/mol, e se trata de uma reação endergónica em 10.52 kcal/mol. Durante o segundo passo, o intermediário quinonóide é protonado pela Tyr-334B, formando-se a histamina que é o produto final da reação. Este passo é consideravelmente mais rápido que o primeiro ($E_a$ = 2,87 kcal/mol) e, em contraste, é altamente exotérmico ($E_r$ = -26,02 kcal/mol).

Este trabalho permitiu também obter uma descrição atomística do mecanismo catalítico da HDC, revelando quais os resíduos do centro ativo que têm um papel determinante para a catálise. Estes resultados podem agora ser úteis para o desenvolvimento de novos fármacos que possam ser usados na terapêutica de doenças relacionadas com a produção de histamina. Além disso, este trabalho corrobora a importância dos métodos computacionais, uma vez que se tratam de uma excelente ferramenta para complementar os dados experimentais e que permitem também muitas vezes corrigir conceções incorretas, promovidas por estes, tal como aconteceu neste caso em particular.

**Palavras-Chave**

Histidina descarboxilase, piridoxal-5'-fosfato (PLP), *L*-histidina, histamina, mecanismo catalítico, mecânica quântica, métodos híbridos QM/MM, ONIOM, catálise enzimática, Teoria do Funcional de Densidade (DFT)

# Preamble

This dissertation is composed of four major chapters and four appendices:

**A. Introduction**

A brief description of Histidine decarboxylase and PLP-dependent enzymes is presented. In this chapter, it will be described some available details about the catalytic mechanism, the gene regulation, the protein maturation, mutagenesis studies, test of inhibitors and biological relevance of Histidine decarboxylase. Additionally, a brief description of PLP-dependent enzymes, and PLP as cofactor will be included.

**B. Theoretical Background**

A short review of the theoretical methods used during this project is described. Particular attention is given to molecular mechanics, quantum mechanics, and hybrid methodologies.

**C. Results and Discussion**

The obtained results will be chronologically presented in this chapter, as well as, some insights about the methodologies that were applied along the work that was done. The first section is about the results associated to the molecular dynamic simulations. Then, the QM/MM results will be presented, as well as, all the required modification to the PDB file, aiming to model, as accurately as possible, the catalytic mechanism of mHDC.

**D. Conclusions and Future Perspectives**

This final chapter will present the major conclusions of this work, and the relevant contributions for the advances in the knowledge provided about Histidine decarboxylase.

**E. Bibliography**

List of all bibliographic references used along the dissertation. They are sorted in the order they appeared in the text.

## F. Appendices

Four appendices were attached to this dissertation:

1. Detailed Information about the molecular dynamics simulations

In this section, a more detailed information about the methods and the structure of Histidine decarboxylase will be presented resorting to the molecular dynamic simulation.

2. Book Chapter

A book chapter about PLP-dependent enzymes and their importance in the cancer treatment was written during this project. The book chapter, entitled “Cancer therapies based on Enzymatic Amino-Acid Depletion”, was already accepted for publication in the multi-volume SET (I-V) “Therapeutic Nanostructures” - Elsevier.

3. Conference Presentations

This section includes all the contributions of this work to conference presentations (posters and oral communications).

4. Non-academic work

During this project, an education game was developed in order to explain, in a playful way, the importance of the drug development in the treatment of different diseases. Furthermore, the TCL scripts that were developed to handle with Gaussian input and output files will be described.

# Table of contents

# Index of figures

# Index of equations

This page was intentionally left blank.

# Index of tables

This page was intentionally left blank.

# List of abbreviations

| | |
|---|---|
| 5-HTP | 5-hydroxytryptophan |
| AADC | Aromatic-*L*-amino acid decarboxylase |
| AANM | Age at natural menopause |
| AM1 | Austin Model 1 |
| AMBER | Assisted Model Building with Energy Refinement |
| B3 | Becke 3 Functional |
| B3LYP | Becke 3 and Lee, Yang and Parr Functional |
| BJ | Becke and Johnson Functional |
| CAM | Coulomb-Attenuating Method |
| cAMP | Cyclic Adenosine Ponophosphate |
| cdk | Cyclin-dependent kinase |
| cDNA | Complementary Deoxyribonucleic acid |
| CDX1 | Caudal-type homeodomain transcription factor |
| CGTO | Contracted Gaussian-type orbitals |
| CHARMM | Chemistry at Harvard Macromolecular Mechanics |
| CHF | Chronic heart failure |
| CNDO | Complete Neglect of Differential Overlap |
| CNS | Central Nervous System |
| compASM | Computational alanine-scanning mutagenesis |
| CPK | Ball-and-stick representation |
| DAG | Diacyl glycerol |
| DAO | Diamine oxidase |
| DCF | Dispersion Correction Functional |
| DFT | Density-Functional Theory |
| DNA | Deoxyribonucleic acid |
| EA | External aldimine |
| EA-P | External aldimine with the product |
| EA-S | External aldimine with the substrate |
| ECL | Enterochromaffin-like |
| EGCG | epigallocatechin-3-gallate |
| ER | Endoplasmic Reticulum |
| ERK | Extracellular signal-regulated kinase |
| FMH | S-(α)-fluoromethyl-histidine |
| FT97 | Filatov-Thiel 1997 Functional |

| | |
|---|---|
| FXR | Farnesoid X receptor |
| GAD | Glutamate decarboxylase |
| GAFF | General AMBER force field |
| GGA | Generalized Gradient Approximation |
| GROMOS | Groningen Molecular Simulation package |
| GTO | Gaussian-type orbitals |
| HDC | Histidine decarboxylase |
| HF | Hartree-Fock |
| HFT | Hartree-Fock Theory |
| hHDC | Human Histidine decarboxylase |
| HL | High-level |
| HME | Histidine methyl ester |
| HNMT | Histamine N-methyltransferases |
| IA | Internal aldimine |
| Ig | Immunoglobulin |
| INDO | Intermediate Neglect of Differential Overlap |
| $IP_3$ | Inositol 1,4,5-triphosphate |
| KO | Knockout |
| L-DOPA | L-3,4-dihydroxyphenylalanine |
| LCAO | Linear Combination of Atomic Orbitals |
| LDA | Local Density Approximation |
| LG | Lacks and Gordon Functional |
| LL | Low-level |
| LSDA | Local Spin-Density Approximation |
| LYP | Lee, Yang and Parr Functional |
| M05 | Minnesota 05 Functional |
| M05-2X | Minnesota 05-2X Functional |
| M06 | Minnesota 06 Functional |
| M06-2X | Minnesota 06-2X Functional |
| McN-A-1293 | 4-imidazolyl-3-amino-2-butanone |
| MD | Molecular Dynamics |
| ME | 3-methoxy-5,7,3',4'-tetrahydroxyflavan |
| MEK | ERK kinase |
| mHDC | Mammalian Histidine decarboxylase |
| MM | Molecular Mechanics |
| MNDO | Modified Neglect of Diatomic Overlap |

| | |
|---|---|
| Mr | Relative molecular mass |
| mRNA | Messenger Ribonucleic acid |
| NDDO | Neglect of Diatomic Differential Overlap |
| NF-κB | Nuclear factor κ-light-chain-enhancer of activated B cells |
| NREM | Non-rapid eye movement sleep |
| NSD-1055 | 4-bromo-3-hydroxybenzyloxyamine |
| ONIOM | Our own N-layered integrated molecular orbital molecular mechanics |
| OPLS | Optimized Potentials for Liquid Simulations |
| P86 | Perdew 86 Functional |
| PBE | Perdew-Burke-Ernzerhof Functional |
| PDB | Protein Data Bank |
| PES | Potential Energy Surface |
| PEST | Proline, glutamic acid, serine and threonine containing sequence |
| PLC | Phospholipase C |
| PLP | Pyridoxal-5'-phosphate |
| PM3 | Parameterized Model 3 |
| PM6 | Parameterized Model 6 |
| Pos-HF | Pos-Hartree-Fock |
| PW | Perdew-Wang Functional |
| PW91 | Perdew-Wang 91 Functional |
| QI | Quinonoid intermediate |
| QM | Quantum Mechanics |
| RDF | Radial Distribution Function |
| REM | Rapid eye movement sleep |
| RMSd | Root-Mean-Square deviation |
| RMSF | Root-Mean-Square Fluctuations |
| SCF | Self-consistent field |
| SHP | Small heterodimer partner |
| STO | Slater-type orbitals |
| TS | Transition state |
| VDW | Wan der Waals representation |
| VMD | Visual Molecular Dynamics software |
| vs | Versus |
| ZPE | Zero-point energy |

This page was intentionally left blank.

# A. Introduction

*It is health that is real wealth and not*
*pieces of gold and silver.*
**Mahatma Gandhi**

This page was intentionally left blank.

# 1. Histamine

Histamine is a very important biological amine (**Fig. 1**) that plays an important role in a variety of reactions *in vivo* such as: dilatation in the cardiovascular system, contraction of smooth muscle, regulation of gastric acid secretion [1, 2] and a neurotransmitter in the nervous system [3, 4]. Moreover, histamine is also important in the immune response against pathogens, where it triggers an inflammatory process that increases the capillary permeability to leukocytes. [3, 4]

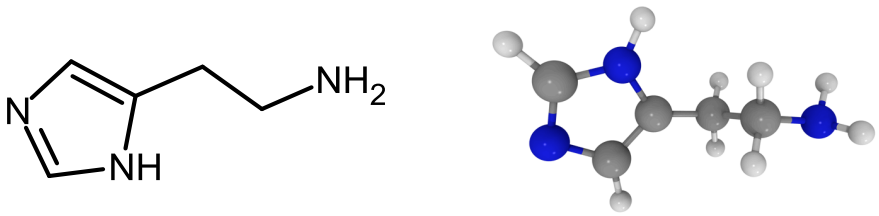



**Fig. 1 – Structure of histamine.**

In mammals, histamine is predominantly produced in the immune system by mast cells and basophils. It is also produced in the central nervous system (CNS) by the tuberomammillary nucleus of the posterior hypothalamus, and in the stomach by enterochromaffin-like cells (ECL). Once produced, it is often stored in fast or slow renewal granules. The slower process is generally associated with immune response events, where histamine is slowly stored in secretory granules of mast cells and basophils. [5, 6] The slower storage prevents exuberant and exaggerated inflammatory responses that could injure the tissues.

The fast renewal occurs in the granules of gastric cells and mainly in the neurons because they are frequently stimulated and require a higher supply of histamine. [5, 6] Contrarily to what happens in the immune system, the exaggerated histamine release in the CNS is not a problem since, once it is released, it is limited to the synaptic cleft where it is rapidly degraded.

The release of the stored histamine occurs through immunological dependent or independent processes. The associated immune process is the most relevant one, and it is triggered by the antigen-IgE (Immunoglobulin E) interaction, leading to a mast cell or basophil degranulation. The independent process, i.e., from a mechanical or chemical stimulus, occurs through the degranulation of histaminergic neurons and ECL cells. [7]

Histamine can also be produced and be rapidly released from the cell in response to an extracellular stimulus. This histamine, also called induced or nascent histamine, plays an important role in the cell growth [2], in the platelet aggregation [8], and in the

gastric acid secretion [2, 9]. Furthermore, this type of histamine is believed to act soon after its synthesis at the site or in the cell where it is formed. [2]

Once released (**Fig. 2**a), histamine has a limited time of action. It is rapidly degraded by histamine N-methyltransferases (HNMT) and diamine oxidase (DAO) (**Fig. 2**b) into methylhistamine and imidazole-4-acetic acid, respectively, whose metabolic products are excreted in the urine (**Fig. 2**c). [7] Interestingly, histamine release is accompanied by an autacoid effect that not only affects the cell that produces it but also the adjacent ones.

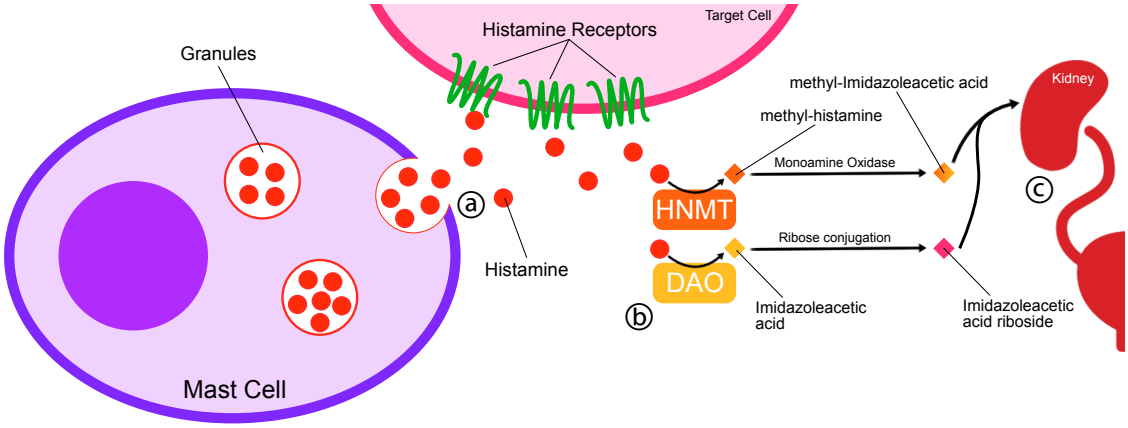



**Fig. 2 – Histamine release, receptors activation, metabolism, and excretion.**

### 1.1. Histamine Receptors

Once histamine is released, it triggers a variety of receptors on the neighbor cells. There are four G-protein-coupled receptors in the cells, called $H_1$, $H_2$, $H_3$, and $H_4$, that are activated depending on histamine availability. The $H_1$ receptor is associated with a $G_{q/11}$ protein, which activates the phospholipase C (PLC), increasing the intracellular concentration of inositol 1,4,5-triphosphate ($IP_3$) and diacylglycerol (DAG). The $IP_3$ binds to the calcium channels in the endoplasmic reticulum (ER), increasing by this way the calcium concentration in the cytosol. [10] Consequently, the NF-κB transcription factor is activated and induces the expression of immune response-related genes [11] and also nervous system plasticity related with memory and learning [12]. The consequences of this cascade of events, triggered by the $H_1$ receptors activation, are bronchoconstriction, capillaries dilatation, an increase of the vascular permeability and afferent neurons sensitization. [7]

The $H_2$ receptor triggers a $G_s$ protein that activates adenylyl cyclase and thereof increases the intracellular concentration of cAMP. This process activates several other

enzymes that stimulate the gastric acid secretion, the cytokines production, the T-cell proliferation and vasodilatation. [7]

Both $H_3$ and $H_4$ receptors activate a $G_i$ protein that decreases the intracellular concentration of cAMP. This means that their activation inhibits the effects triggered by the $H_2$ receptor. [7] Among these two receptors, the $H_3$ receptor plays a major role in the CNS since it is essential to control the circadian rhythms. This receptor initiates a negative feedback mechanism forcing histamine release reduction. The functional role of the $H_4$ receptor is still not clear, although its expression has been reported to be specifically associated with the human bone marrow and colon and it could also be involved in the immune response. [13] In contrast, the $H_1$, $H_2$, and $H_3$ receptors are ubiquitous in different tissues, including the CNS. [13, 14]

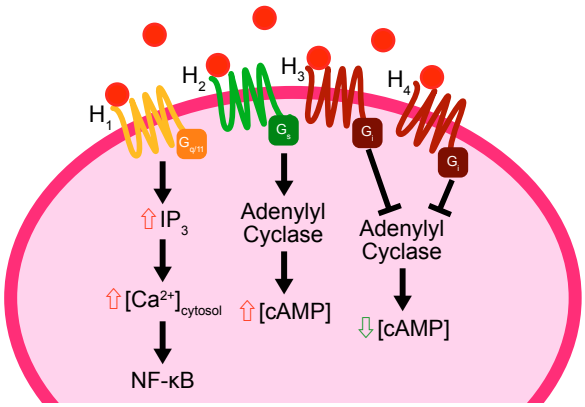



**Fig. 3 – Overview of intracellular pathways triggered by histamine receptors. Red circles represent histamine molecules in the extracellular environment.**

## 2. Histidine Decarboxylase (HDC)

The production of histamine occurs by decarboxylation of *L*-histidine, a reaction that is catalyzed by the Histidine decarboxylase (HDC) (**Fig. 4**) [7, 15, 16]. This enzyme catalyzes the reaction with a rate of 1.73 ± 0.07 $s^{-1}$ and has an enzymatic efficiency ($k_{cat}/K_m$) of 17.3 ± 1.9 $s^{-1}$ $mM^{-1}$. [17]

HDC activity is dependent on the presence of a cofactor in the active site of the enzyme that in mammals (mHDC) is the pyridoxal-5'-phosphate (PLP) and pyruvic acid in bacteria (**Fig. 5**). [18-21] Since in this work we are going to be focused on the mHDC, only PLP-dependent enzymes will be described.

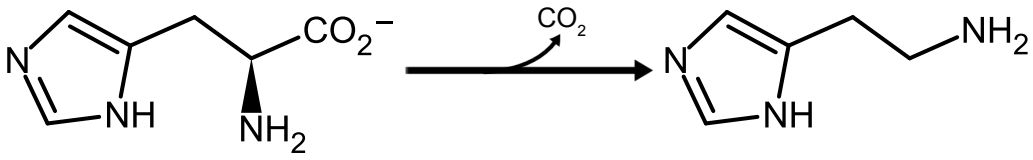
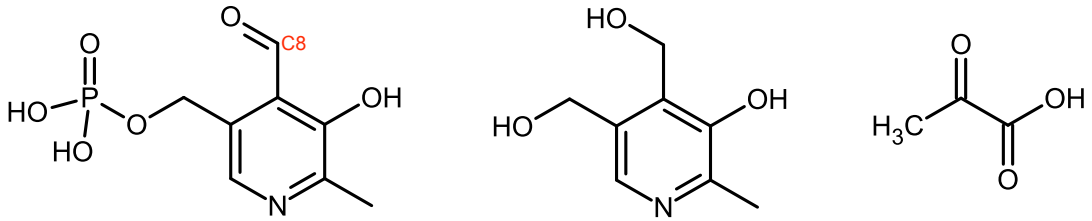



**Fig. 4 - Reaction catalyzed by HDC.**

PLP is the active form of the vitamin $B_6$ (**Fig. 5**), and it is used as a cofactor by more than 140 distinct enzymes, corresponding to more than 4% of all classified enzymatic activities. [22] PLP-dependent enzymes play a crucial role in several metabolic pathways involving amino acids, and catalyze a variety of different chemical reactions, including α-decarboxylation, transamination, aldol cleavage, racemization, β- and γ-eliminations and replacement reactions. [23, 24]




Pyridoxal-5'-phosphate (PLP) Vitamin B6 Pyruvic acid

**Fig. 5 - Wedge-Dash representations of the PLP, vitamin B6, and pyruvic acid.**

Among the PLP-dependent decarboxylases, four evolutionary origins are distinguished in four different families, according to their divergent evolution. Family I includes the glycine decarboxylase that is a component of a multi-enzyme system. [25] Family II includes mHDC (EC 4.1.1.22), L-DOPA (L-3,4-dihydroxyphenylalanine) decarboxylase, also known as Aromatic-*L*-amino-acid decarboxylase, (AADC) (EC 4.1.1.28) and the glutamate decarboxylase (GAD) (EC 4.1.1.15). [26] Family III only contemplates prokaryotic decarboxylases, namely the ornithine and lysine decarboxylase likewise the biodegradative arginine decarboxylase. Family IV includes the eukaryotic arginine and ornithine decarboxylases as well as the prokaryotic biosynthetic arginine and diaminopimelate decarboxylases. [25]

In this work, we will only be focused on family II of the PLP-dependent decarboxylase, because it is where mHDC belongs to. From the enzymes present in this family, AADC is the most promiscuous enzyme because it can catalyze the decarboxylation of *L*-DOPA, 5-HTP (5-hydroxytryptophan), *L*-phenylalanine, *L*-histidine, *L*-tyrosine, and *L*-tryptophan. Both, the mHDC and the glutamate decarboxylase are

highly specific since they only catalyze the decarboxylation of *L*-histidine and *L*-glutamate, respectively. Moreover, animal models showed that mHDC was the main enzyme responsible for histamine biosynthesis *in vivo*. [27]

Apart from the different biological roles assigned to these decarboxylases and a lower homology between HDC and other enzymes from family II (**Table 1**), the structure of the active site is interestingly similar among these enzymes (**Fig. 6**). The major disparities in their sequences are located in the region of the binding site that interacts with the imidazole group of the substrate (**Fig. 6** Right Bottom), so those differences should be correlated with their specificity. **Fig. 6** only shows Tyr-334B and Leu-335B from HDC without the correspondent amino acid residues from AADC and GAD. Indeed, this tyrosine residue is also present in AADC and GAD, but the its-containing loop was not crystallized for AADC, and it is displaced from the active site in the GAD. GAD has the tyrosine-containing loop away because the PDB file presents a structure with the internal aldimine (IA), and it was described that the flexible loop only approximates to the active site when the external aldimine (EA) is formed. This information also explains the displacement of Leu-435B from GAD, since it belongs to the same flexible loop.

In sum, this evidence indicates a possible common catalytic mechanism for all PLP-dependent enzymes from family II, pointing the structural differences of the active site as specificity regulators.

**Table 1– Summary of all PDB structures available for family II of PLP-dependent decarboxylases from mammals. Homology percentage of AADC and GAD comparatively to HDC (human sequences).**

| | Structures | | | | | Homology[3] |
|---|---|---|---|---|---|---|
| Enzymes | Apoenzyme (open) | Apoenzyme (close) | IA | Gem-diamine | EA-Inhibitor | |
| HDC | - | - | - | - | 4E1O[17] | N.A. |
| AADC | 3RBF[28] 3RBL[28] | 3RCH[28] | 3RCH[28] 1JS6[1][29] | - | 1JS3[1][29] | 52% |
| GAD | - | - | 3VP6[30] 2OKJ[31] 2OKK[2][31] | 2OKJ[31] | - | 29% |

[1] No structural information about the loop (amino acid residues: 327 to 338)
[2] Containing only one subunit
[3] Homology was determined using the amino acid sequence of HDC as a reference.

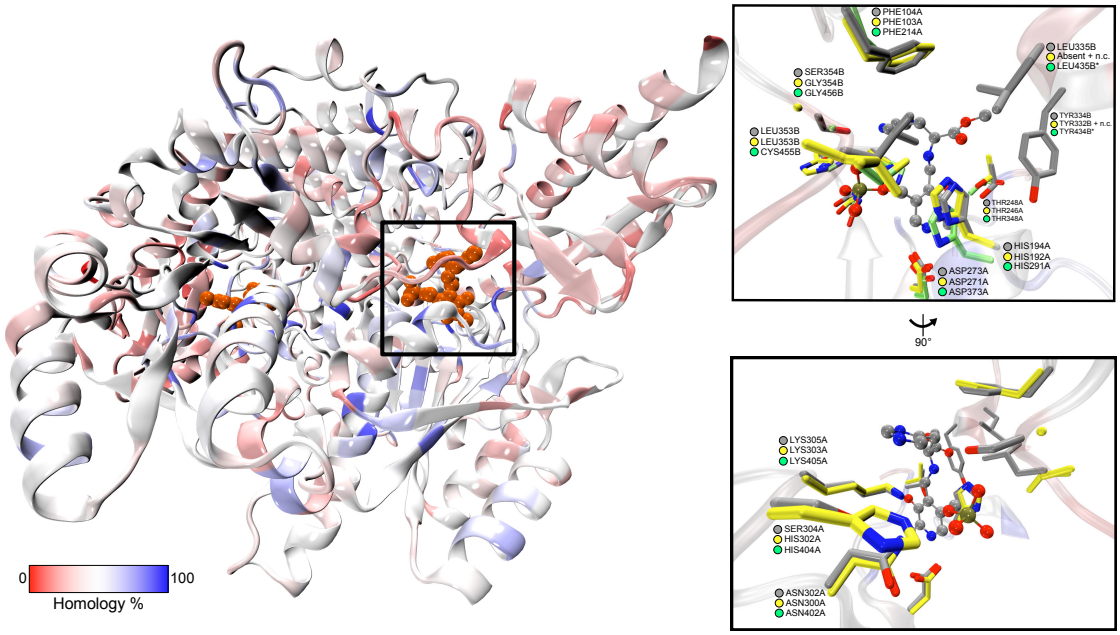



**Fig. 6 – Structural comparison between all the PLP-dependent decarboxylases from family II.**
(Left) NewCartoon representation of HDC colored by homology with AADC. VDW representation of the external aldimine with an inhibitor (histidine methyl ester). (Right) Overlap of the active site from HDC (gray), AADC (yellow) and GAD (green). CPK representation of the external aldimine with the inhibitor. *Amino acid residues are not showed because they belong to a flexible loop which was far away from the active site.

Taking into account all the available crystallographic structures of the PLP-dependent decarboxylases from family II, the only one available for mHDC is the best one. It contains the complete structure of the enzyme co-crystallized with the external aldimine that involves an imine linkage between the PLP cofactor and an inhibitor (Histidine methyl ester - HME). Moreover, HME is a methylated derivative of L-histidine (natural substrate), which helps the construction of the model system of the wild-type enzyme with the natural substrate.

### 2.1. mHDC Gene

Structural analysis of the HDC gene revealed that the human gene is composed of 12 exons spanning approximately 24 kb (kilobases) [32] and located on chromosome 15 [33]. Moreover, DNA analysis revealed a unique copy gene for HDC. Exon 1 contains the 5'-untranslated region, the translation initiation codon, and the coding region for the first ten amino acid residues of HDC, whereas almost half of the enzyme mRNA sequence is encoded by the exon 12, including the 248 carboxyl-terminal amino acid residues. The heterogeneity of the isoelectric point of the enzyme might be due to post-translation modification of the enzyme. [32] Four copies of the GATA consensus sequence were found in the promoter region of HDC gene. [32] Also four GATA-binding

proteins (GATA-1, -2, -3, and -4) have been characterized [34, 35], and GATA binding proteins were found to be expressed abundantly in several mast cell lines through an electrophoretic gel mobility shift assay. [36]

cDNA analysis revealed that the calculated molecular weights of the rat, mouse, and human histidine decarboxylases (hHDC) deduced from the nucleotide sequences are significantly larger than those determined for the purified proteins, suggesting that primarily translated products are post-translationally processed to yield mature active enzymes. [32]

Expression of HDC is regulated by several different *stimuli*, including gastrin [37], phorbol ester 12-myristate-13-acetate [37], oxidative stress [38], thrombopoietin [39], and *Helicobacter pylori* infection [40].

Cell stimulation with either gastrin or phorbol ester 12-myristate-13-acetate is known to increase the phosphorylation of extracellular signal-regulated kinases (ERKs) and increase ERK activity. Not unexpectedly, the overexpression of Erk-1 and Erk-2, or activation of endogenous ERKs using activated MEK-1 (ERK kinase-1), stimulate HDC-promoter activity. Additionally, gastrin regulates HDC transcription in a Raf-dependent, Ras-independent fashion through activation of the ERK-related pathway. [13]

HDC gene expression varies with time during the day and selectively deleting the Bmal1 clock gene from histaminergic neurons removes this variation, producing higher HDC expression and brain histamine levels during the day. The consequences include more fragmented sleep, prolonged wake at night, shallower sleep depth, increased NREM-to-REM (Non-rapid eye movement sleep to rapid eye movement sleep) transitions, hindered recovery sleep after sleep deprivation, and impaired memory. The local BMAL1-dependent clock directs appropriately timed declines and increases in histamine biosynthesis, by HDC gene expression, to produce an appropriate balance of wake and sleep within the overall daily cycle of rest and activity specified by the suprachiasmatic nucleus. [41, 42]

Also, epigenetic patterns were described as been an important regulator of HDC expression, once DNA demethylation by 5-azacytidine induces high expression of HDC mRNA. [13]

### 2.2. mHDC Structure

The structure and function of mHDC are the less described in the literature comparing with the other decarboxylases from family II, such as AADC. Nonetheless, some works have reported certain similarities between some of mHDC and AADC motifs. [43, 44]

mHDC is firstly translated into an insoluble form with a size of 74 kDa (**Fig. 7**a), but a 54 kDa form has also been identified. This size difference is caused by a post-translational process (**Fig. 7**b) induced by a sequence of amino acids denominated as PEST regions that are rich in proline, glutamic acid, serine, and threonine. The PEST sequences are intrinsically associated with a short half-life time of proteins, and they act as peptide signals for protein degradation. [45, 46] In the case of mHDC, it has two PEST regions. One is located near the *N*-terminus (between the amino acid residues 40 and 70) and another in the *C*-terminus region, after the amino acid residue 503. [43] The *C*-terminus PEST region triggers the proteolytic activity of this region promoting the monomer maturation (**Fig. 7**b). This proteolytic maturation converts the 74 kDa form into the 54 kDa monomer. [43] After this maturation, another PEST sequence remains in the protein indicating a short half-life time for this protein, once it can initiate the protein degradation (**Fig. 7**c). In fact, the 54 kDa form of the mHDC indeed has a short half-life time of 50 to 100 minutes. [26]

Both forms of the enzyme possess the same histamine synthesis capability, and a hypothesis was described to explain the existence of both forms. Yatsunami's group proposed that the 74 kDa form bind to the cell membrane and promptly responds to signals from the extracellular environment, whereas the 54 kDa form is dispersed along the cytosol. [47] The Yatsunami's hypothesis indicates that the 74 kDa form is responsible for the newly formed histamine, once it is suitable to respond easily to an extracellular stimulus (**Fig. 7**d). On the other hand, the 54 kDa form is a soluble form that produces the stored histamine (**Fig. 7**e). [47]

Subsequent studies in RBL-2H3 (basophils cell line) cells demonstrated that the degradation of the 74 kDa form of mHDC is an ATP-dependent process, under the ubiquitin-conjugated pathway, and it is sensitive to proteasome inhibitors (**Fig. 7**b). These results, together with the significant activity of proteasome in RBL-2H3 cells, demonstrated that the 74 kDa form is a short half-life protein because of its rapid degradation via the proteasome pathway (**Fig. 7**b). When compared to the 74 kDa form, the 54 kDa form seems to be relatively stable. In contrast with previous studies [47], the 74 kDa form is localized primarily in the cytosol, whereas the 54 kDa form is mostly located in the endoplasmic reticulum, Golgi apparatus, and granules. [48, 49] Tanaka's

work completes Yatsunami's hypothesis once it revealed that the 74 kDa form may have a function in the production of cytosolic histamine that is spontaneously released (**Fig. 7**d) or transported to histamine-containing granules for storage (**Fig. 7**f) Whereas the 54 kDa form is convenient to synthesize directly granular histamine (**Fig. 7**e), which is stored and released in response, for example, to a receptor-mediated immune reaction (**Fig. 7**g). [48]

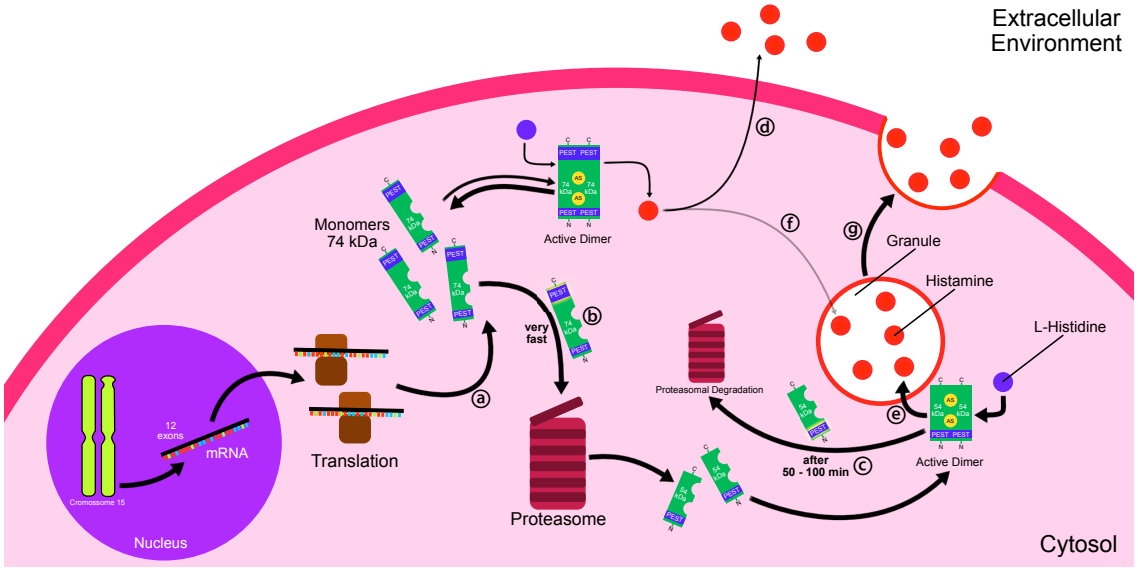



**Fig. 7 – Schematic representation of the mHDC synthesis and post-translational processing.**

The first and, until now, unique experimental mHDC structure was obtained in 2012 by Yoshiki Higuchi's research group. It was necessary to make a double mutation at C179S and C417S (PDB equivalent to C180S and C418S) to prevent nonspecific protein oligomerization, which was found to turn difficult the structure determination by X-ray crystallography. [50] The wild-type enzyme and the double mutated form showed similar $K_m$ and $k_{cat}$ values, indicating that the mutation of these two cysteine residues on the protein sequence did not affect enzyme activity. [17] Structure information was deposited into the Protein Data Bank with the PDB code 4E1O, and contains three dimers with the PLP cofactor and a histidine derivative inhibitor, the histidine methyl ester (HME), bounded in the active site of each monomer. The two active sites are placed in the interface of the two subunits, and they are composed of amino acid residues from both monomers (**Fig. 8**). [50]

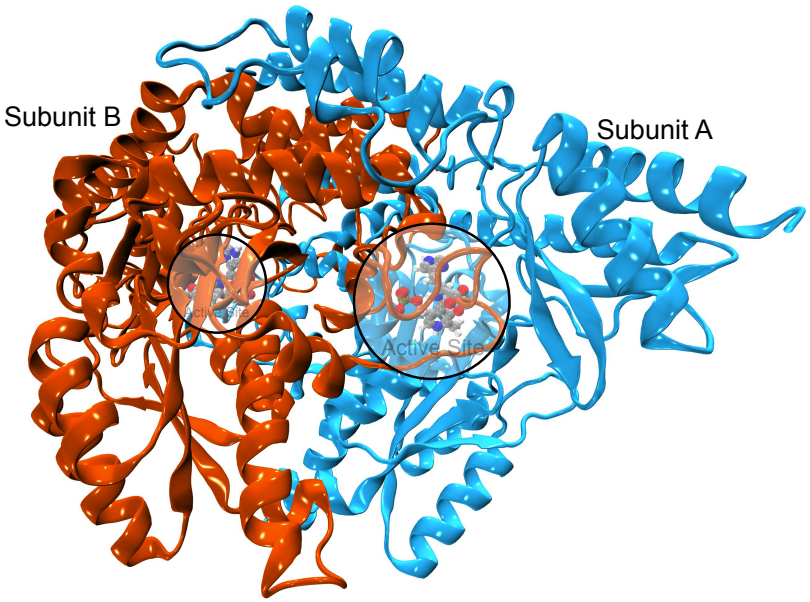



**Fig. 8 – Structure of mHDC dimer and the external aldimine of each active site.**

Each subunit is composed of 477 amino acid residues comprising the post-translational form of the mHDC (54 kDa form). The subunit could be divided into three structural domains: an *N*-terminal (2–71), a large (71–371) and a small (372– 477) domains. The large domain is connected to the small domain by a long α-helix. The dimer interface is formed by the *N*-terminal and large domains, producing a stable dimeric structure. The *N*-terminal regions of each monomer are mainly hydrophobic and are located at the dimer interface. The large domains are exposed to the solvent and point to the opposite direction of the dimer interface. [17]

## 2.3. Catalytic Mechanism

The catalytic mechanism of mHDC can be divided into three main stages: the PLP activation, the transimination, and the decarboxylation. The first two stages are common among all the PLP-dependent enzymes, and only the last one is specific to each enzyme.

### 2.3.1. PLP Activation and Transimination Mechanisms

All known PLP-dependent enzymes follow a common set of initial steps which activates the enzyme through the formation of an internal aldimine (IA). In such state, the PLP cofactor becomes bonded to an active site lysine by a covalent bond (**Fig. 9**).

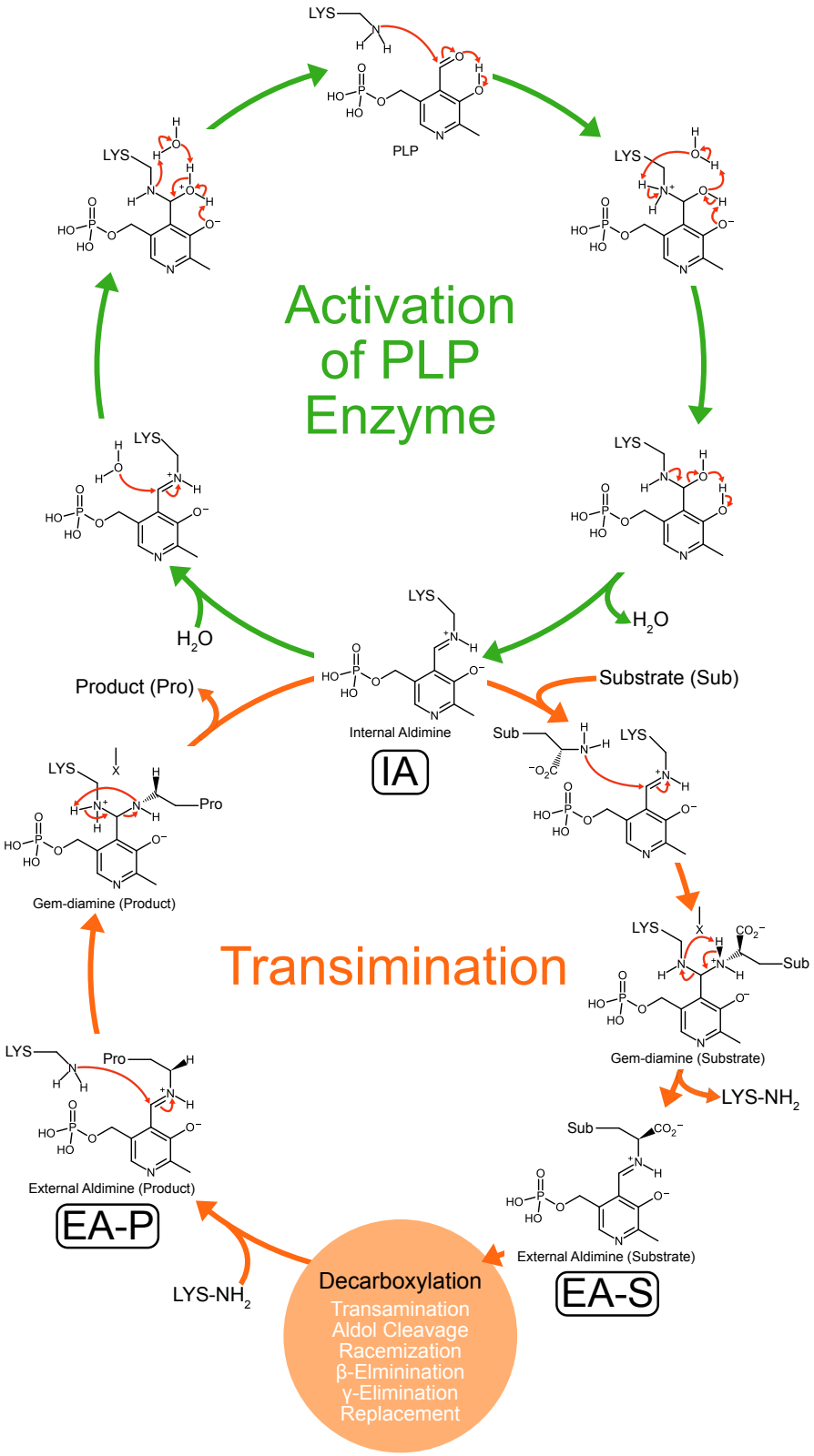



**Fig. 9 – Schematic representation of the currently accepted mechanism for the internal and external aldimines formation. [51, 52]**

The fixation of the cofactor to the enzyme is mediated by an imine bond that is established between the ε-amino group of the lysine residue and carbon C8 of PLP (**Fig. 5**). In this process, one water molecule is released. [52, 53]

Once the internal aldimine is formed, the enzyme is activated, and the decarboxylation process can take place. If the substrate is available (for example histidine), it binds to internal aldimine and forms a gem-diamine structure. Then, one proton from the imine group of the substrate migrates to the imine group of the lysine leading to the dissociation of the active site lysine. At the end of this process, the substrate becomes bonded to the PLP cofactor through an imine bond. This intermediate is often called external aldimine (EA), but as in the PLP mechanisms, there are several different external aldimine species we will call it EA-S to denote that it is an external aldimine with the substrate bonded to it. The EA-S is then ready to undergo a set of different reactions as mentioned before. Those reactions are the ones that make the PLP-dependent enzymes specific for different reactions. In the case of mHDC, it is a decarboxylation process. Once the specific reaction takes place, the external aldimine with the product of the reaction (EA-P) is formed, and the enzyme is ready for a new turnover, once the product of the reactions dissociates from the PLP cofactor and the active site lysine becomes bonded to PLP (IA) (**Fig. 9**). [51]

In the case of mHDC, these mechanistic steps that were studied by computational means [51, 52] are confirmed by spectrophotometric studies. In the wild-type enzyme, the internal aldimine presents a characteristic absorption spectrum with two maximum peaks around 335 and 425 nm (**Fig. 10** – Blue line), which are generally attributed to the enolimine and ketoenamine tautomeric forms (**Fig. 11**) of the IA, respectively. The absorption peak at 335 nm is twice as high as that at 425 nm (**Fig. 10** – Blue line), indicating a prevalence of the enol form. [54] The predominance of the enol form of the IA is considered to reflect the hydrophobic nature of its active site that accepts aromatic substrates. [54] This trend was also observed for the AADC, which also catalyzes the decarboxylation of aromatic substrates. [55]

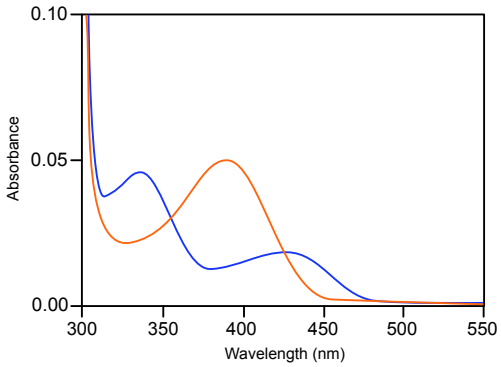



**Fig. 10 – Absorption spectra of HDC in the absence (blue line) and presence (orange line) of 1 mM HME (histidine methyl ester).**
The spectra were taken in 50 mM PIPES-NaOH buffer, pH 7, at 25 °C. The enzyme concentration was 8 μM. [54] (Adapted)

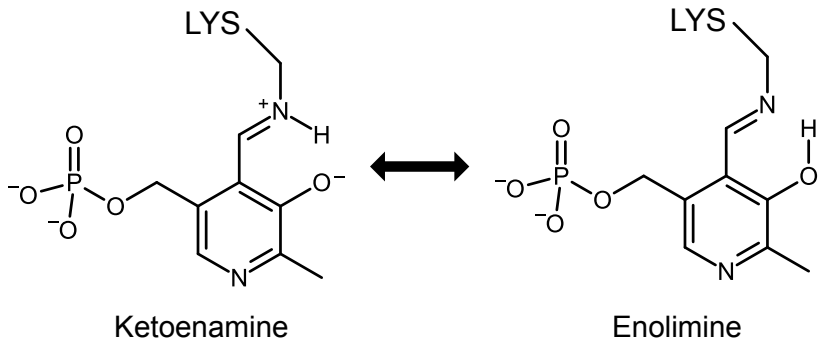



**Fig. 11 - Wedge-Dash representations of the internal aldimine in a (Left) ketoenamine form and (Right) enolimine form.**

These initial steps of the mechanism are also confirmed by a spectroscopic study of the wild-type enzyme in the presence of an inhibitor (Histidine methyl ester - HME). In this case, it is observed a decrease in the absorbance at both 335 and 425 nm, and a significant increase of the absorbance around 388 nm (**Fig. 10**). [54] This absorption band indicates the accumulation of the EA-S that stalls the reaction in this reaction intermediate. Similar results have also been observed for rat AADC with DOPA methyl ester (similar inhibitor). [55] The same experiment have also detected a low at 425 nm, corresponding to the keto form of the EA-S. [54]

### 2.3.2. mHDC mechanism

Once the EA-S is formed, a certain number of particular chemical reactions happen, and they make the difference between all PLP-dependent enzymes. Concerning about the mHDC, a decarboxylation takes place, and the resultant quinonoid intermediate (QI) is subsequently protonated.

To evaluate, sequentially, the histidine decarboxylation, a stopped-flow spectrophotometer study was done with the natural substrate (*L*-histidine). In the first fast phase, a rapid increase in the absorbance at 425 nm with a concomitant decrease in the 335-nm peak was observed within 50 milliseconds, showing a shift from the IA to the ketoenamine tautomer of the EA with the substrate (EA-S). [54] These spectral changes are also similar to those observed for the reaction of AADC and L-DOPA. [55] In a second slower phase, the absorption at 425 nm decreases, and there is the concomitant emergence of a new absorption band at around 380 nm within 2.0 seconds, indicating that the enolimine tautomer of the EA is formed. However, no distinction between the EA-S and the EA-P (external aldimine with the product) could be made. [54]

Circular dichroism spectroscopy assays have also revealed an inversion of the conformation of the Schiff base. This inversion is probably caused by the change in the amino moiety of the Schiff base from the IA to the EA-S. Comparing the obtained data, the EA of mHDC has a more distorted structure than that of AADC. [54, 55] In this sense, it is interesting that the overall rate constant of the steps after the EA-S formation, which includes the decarboxylation step, is apparently lower for mHDC than for AADC. Thus, we can consider that the distorted conformation of the EA in the mHDC lowers the rate of decarboxylation by partially breaking the π-conjugation between the pyridine ring and the imine. [54]

When the α-fluoromethyl-histidine inhibitor (FMH) is used, a similar QI can be accumulated in a stable complex. The gel filtration of the complex with the similar quinonoid leads to an apparent reduction of the $M_r$, suggesting a “more closed” structure. [56]

Rodriguez’s group performed temperature denaturation assays that showed more evident changes for the free holoenzyme comparatively with the α-FMH-treated enzyme sample, indicating that the enzyme that had an EA is more resistant to temperature-induced denaturation. However, the observed increasing trend in the ellipticity at 222 nm of the HME-treated enzyme preparations suggests that the reversible inhibitor is not able to protect the enzyme against thermal denaturation. [56] Those results indicate that mHDC adopts at least two well-differentiated conformations during the catalytic reaction. One corresponds to the fully active IA of the enzyme, and the

second one takes place in the presence of the external aldimine (EA-S or EA-P) in the catalytic site. [56]

Other earlier studies have already shown that histamine is produced with the retention of configuration in HDC from *Lactobacillus 30a*, a pyruvate-dependent decarboxylase. [57] The assay was reproduced with a PLP-dependent decarboxylase model, and the result was also corroborated for those enzymes. [58] These observations indicate a previous decarboxylation, and a subsequent proton addition at the same position. Therefore, an $S_N2$ nucleophilic attack is impossible to take place once it would add a hydrogen in an *anti*-position relatively to the liberated carboxylate group. Then, the carboxylate group should be released in the first step, and afterward, a proton is added to the α-carbon. The first step (decarboxylation) is promoted by the strongly electron-withdrawing nature of the pyridoxylidene moiety that weakens the bond between the carboxyl group and the extended π-system of the Schiff bases formed with the substrate (*L*-histidine) and PLP, resulting in the loss of the carboxylate group. [53, 59] In 1966, Dunathan has already proposed that all double bonds of the extended π system should be placed on the same plane of the PLP ring to favor the electron displacement and by this way stabilize the QI. During the decarboxylation process, the bond that is broken, the Cα of PLP-amino acid Schiff's bases should, however, be held at 90° to the plane in the relation the extended conjugated system to promote a faster catalysis. [60, 61]

In the second step, the QI is protonated. However, the proton donor is not known. An earlier computational study has suggested that the lysine residue (Lys-305A) could protonate the carbanion formed [62], but recent experimental structural data showed a huge impediment for the protonation process by Lys-305A. Additionally, according to the crystallographic structure, the Lys-305A is on the opposite face of the carboxylate group which does not agree with that computation study. [17] In sum, there are no suggestions about which amino acid residue or molecule can protonate QI at the active site.

### 2.3.3. Mutagenesis studies

mHDC holoenzyme seems to have a highly restrictive and hydrophobic binding pocket. It is only able to bind histidine or imidazole-containing analogs, such as the inhibitors FMH and HME. [26] In the following topic, the active site features will be depicted to look for hints about the structural arrangement and provide insights into the catalytic mechanism of the enzyme.

As mentioned before, the enzyme activation through the IA formation is mediated by a lysine amino acid residue (**Fig. 9**) and some studies with HDC from *Morganella morganii* and other PLP-dependent enzymes revealed that this lysine residue (Lys-305A)

(**Fig. 12**) is essential for the enzymatic activity. The replacement of this lysine by an alanine [63] or an arginine [64] culminates into the full inactivation of the enzyme. The substitution of this residue by an arginine residue allows the IA formation, but it does not allow the progress of the catalytic reaction, presumably stalling the enzyme in the IA. [64, 65]

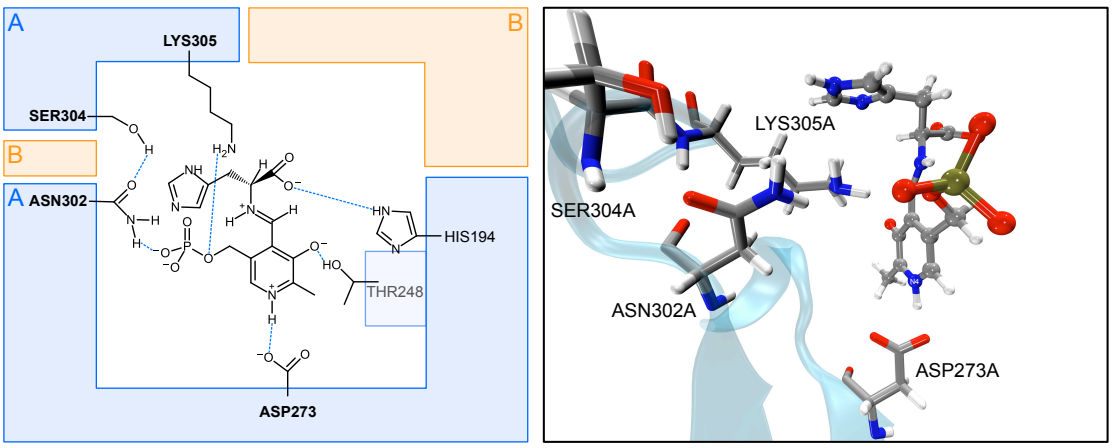



**Fig. 12 – (Left) Wedge-Dash representation of the external aldimine (EA-S) and the Asp-273A, Lys-305A, Ser-304A and Asn-302A residues.**
Blue dashed lines represent relevant non-bonded interactions (Right) CPK and licorice representations of the external aldimine (EA-S) and the Asp-273A, Lys-305A, Ser-304A and Asn-302A residues, respectively.

An extensive experimental mutagenesis study, that tested 27 different mutant proteins, was done by Flemming and co-workers that studied the role of 34 different amino acid residues in the catalysis. These results showed that several amino acid residues are indispensable for catalysis. In this study, the importance of the Lys-305A (**Fig. 12**) was confirmed once its substitution by a glycine totally inactivates the enzyme. [66] Also, the mutation of the Asp-273A (**Fig. 12**), an amino acid residue that interacts with N4 of PLP-pyridine, by glycine also leads to the inactivation of the enzyme. [66]

An almost inactivation of the enzyme is also achieved when Ser-304A (**Fig. 12**) is mutated by a glycine. This residue defines together with the Lys-305A a α-helix that holds the position of Lys-305A that is essential for the IA formation. This serine residue is also involved in a hydrogen bond with the Asn-302A (**Fig. 12**) that helps the stabilization of the phosphate group of the PLP. [66] The same type of inhibition is also achieved when His-194A (**Fig. 13**) is mutated by a glycine. [66] This amino acid residue makes a π stacking interaction with the ring of the PLP, and therefore it is expected that its mutation might lead to a difference rearrangement of the PLP ring and accordingly affect the catalytic process. Additionally, this residue establishes a hydrogen bond with the carboxylate group of the substrate, which may be important for the catalytic process (**Fig. 13**). [17]

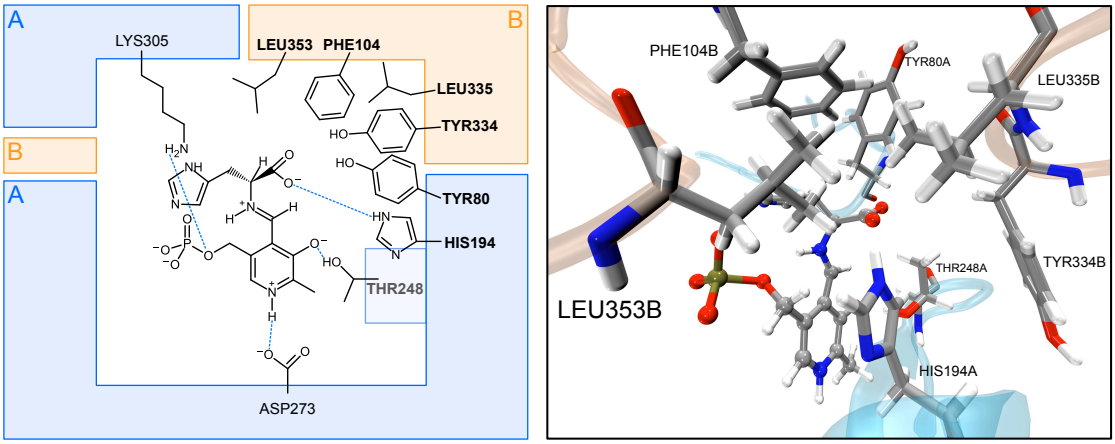



**Fig. 13 - Wedge-Dash representation of the external aldimine (EA-S) and the His-194A, Thr-248A, Tyr-80A, Tyr-334B, Leu-335B, Phe-104B and Leu-353B residues.**

Blue dashed lines represent relevant non-bonded interactions (Right) CPK and licorice representations of the external aldimine (EA-S) and the His-194A, Thr-248A, Tyr-80A, Tyr-334B, Leu-335B, Phe-104B and Leu-353B residues, respectively.

The same mutagenesis studies also indicate that the loop containing Tyr-334B is essential for catalysis (**Fig. 13** and **Fig. 14**). Indeed, the replacement of Tyr-334B or any contiguous amino acid residues of the loop conducts to a total inactivation of the enzyme. This loop (**Fig. 14**) is believed to play a major role in catalysis, as it closes the binding pocket when the substrate is accommodated inside the active site. [66]

Molecular dynamics simulation using a homology model of the mHDC revealed a more noticeable motion of this loop (**Fig. 14**) in HDC with the IA than with the correspondent EA. The IA containing structure is more exposed to the solvent comparing to the structure with the EA. The exposure of the flexible loop to the solvent over the protein surface in IA justifies the more suitability of this portion to be degraded by proteases. [67] The most critical residue in this loop is Tyr-334B, which is also highly conserved in some PLP-dependent enzymes. [29, 68] In a similar test, the equivalent tyrosine residue was replaced by a phenylalanine in the pig AADC, and that substitution changes the reaction that is catalyzed by the enzyme, once decarboxylation-dependent oxidative deaminase happened instead of a decarboxylation followed by the QI protonation. [26] The crystallographic structure obtained in 2012, also indicates that the Tyr-334B may play a critical role in catalysis. [17]

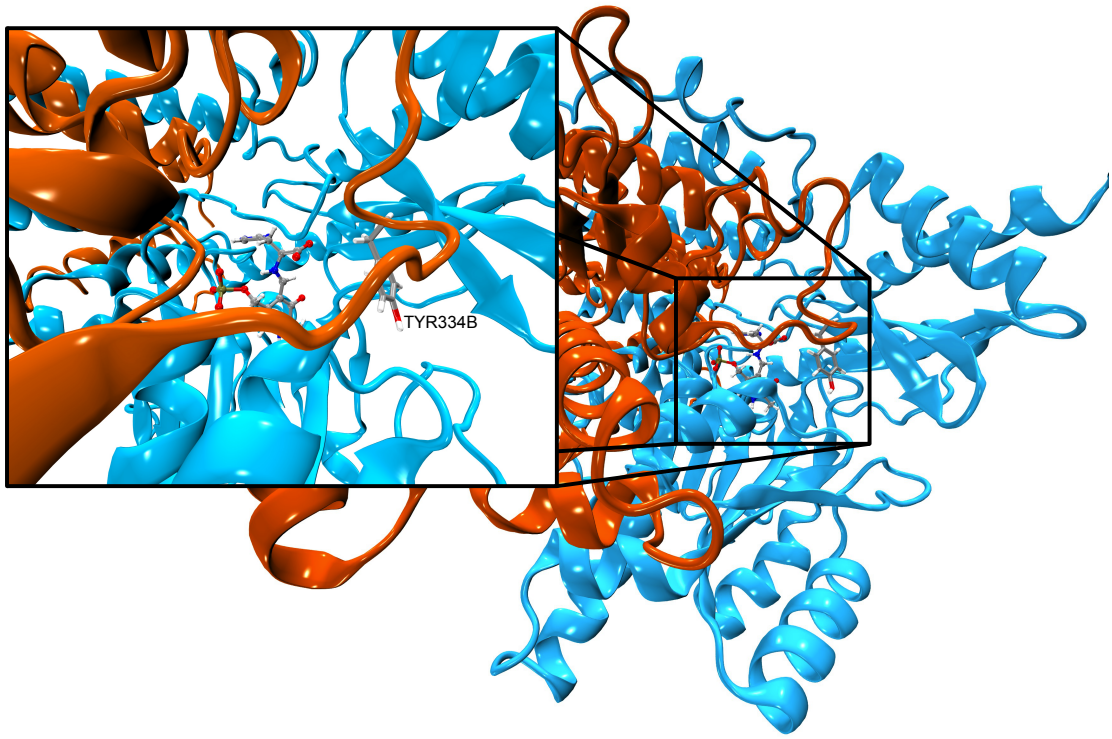



**Fig. 14 – NewCartoon representation of the mHDC dimer.**
The subunit A is blue colored, and the subunit B is orange colored. Zoom at the flexible loop region containing the Tyr-334B (Licorice representation) near the external aldimine with the substrate (EA-S) (CPK representation).

Other relevant mutagenic studies involved the amino acid residues of the active site that establish substrate-specific interactions, i.e., Tyr-83A, Tyr-80A, and Ala-86A. All of these mutations revealed a decrease the enzymatic activity of the mutated enzymes. [26, 66]

Based on the new crystallography structure of mHDC, the previous hypothesis for the structural rearrangement of the PLP cofactor and substrate inside the active site are inaccurate mainly in respect to the *N*-imine orientation and substrate's imidazole group orientation. Previous studies have presented imidazole group pointed to the His-194A face [69], but the new X-ray structure revealed an opposite orientation for the HME's imidazole group. In fact, the imidazole ring of HME points toward the *si*-face of the EA. The nitrogen atoms of the imidazole ring appeared to participate in hydrogen-bonding interactions with the backbone amino group of Tyr-81A and a water molecule. Except for two hydrogen bonds, the imidazole ring was surrounded by hydrophobic residues (Trp-72A, Tyr-80A, Leu-102A, and Phe-104A). Interestingly, flipping the imidazole ring allows the formation of extra hydrogen bonds with Ser-354B (**Fig. 15**) and with the carbonyl group of Tyr-81A, instead of the water molecule and amino group of Tyr-81A, respectively. [17]

The X-ray orientation for the imidazole group was confirmed with a later work that showed an important role of Ser-354B (**Fig. 15**) for substrate specificity for histidine rather than L-DOPA. A mutagenesis study, where the Ser-354B was replaced by a glycine, revealed a 10-fold $K_m$ increase of the S354G mutant when compared with the wild-type enzyme, whereas $k_{cat}$ values for both enzymes were similar, resulting in a decrease of $k_{cat}/Km$ values for the S354G mutant. This data indicates that the mutation of the Ser-354B decreased the affinity between the enzyme and the histidine. Additionally, this mutation increases the binding pocket area nearby the active site which makes mHDC available to catalyze the L-DOPA decarboxylation. [17]

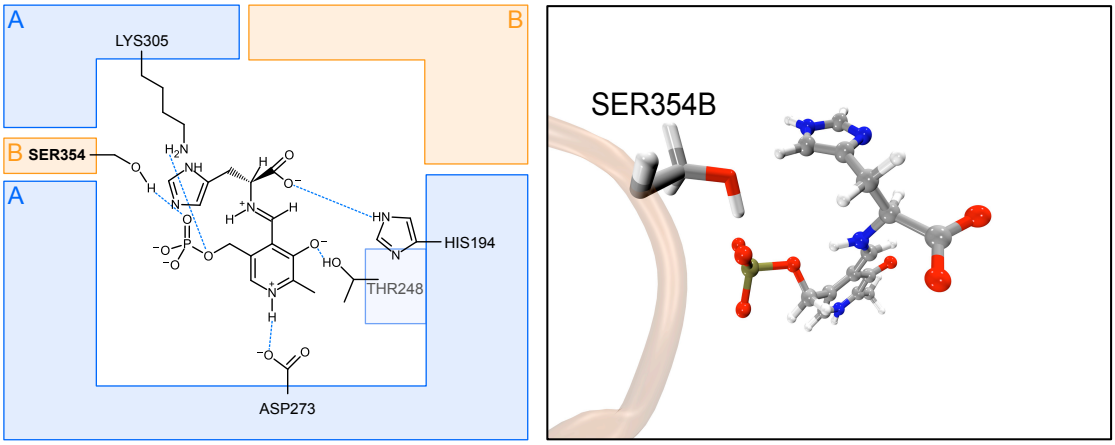



**Fig. 15 – CPK and licorice representations of the external aldimine with the substrate (EA-S) and of the Ser-354B residue, respectively.**

Nevertheless, the *N*-imine group of the EA is pointing to the phosphate group instead of pointing to the enol group of PLP-pyridine as some works have shown before (**Fig. 15**). [26, 54, 56, 62, 69] Additionally, the pyridine ring of the PLP is inserted between the methyl group of the Ala-275A and the imidazole ring of His-194A. The Asp-273A and Thr-248A residues interact, through hydrogen bonds, with the protonated nitrogen and oxygen atoms of the pyridine ring of PLP. The phosphate group of PLP is stabilized by a set of hydrogen bonds with the Ser-151A, Asn-302A, and Ser-354B amino acid residues. Also, the negative charge of the phosphate group interacts with the dipole of helix-5, which was positioned with its *N*-terminus close to the phosphate group. [17]

**Fig. 16** summarizes all important active site residues at the active site of the mHDC, in particular, the red ones, which are preponderant once their mutation completely blocks the reaction that is catalyzed by mHDC.

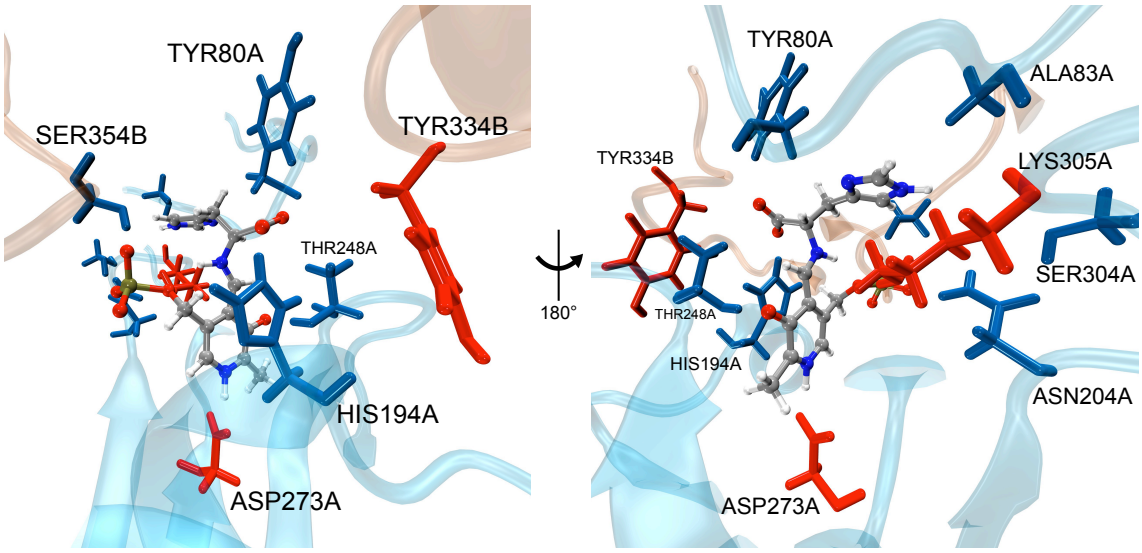



**Fig. 16 – CPK and licorice representations of the EA-S and important amino acid residues at the active site, respectively.**

Red colored amino acid residues are the ones whose mutation completely inactive the mHDC, whereas the blue colored residues only decrease the catalytic rate.

## 2.4. Biological role of histamine in cells

As described before, histamine is widely produced by a variety of cells, and consequently, several physiological conditions are correlated with the histamine release.

### 2.4.1. Cell Cycle

Expression of mammalian HDC (mHDC) induces a cell cycle arrest through the alteration of the levels of several proteins such as cyclin cdk6, D1, cdk7 and cyclin A. Histamine production, mediated by an increased mHDC gene expression and activity, has the ability to regulate caspase-3 and α-synuclein, justifying the relatively reduced proliferation rates of mammalian histamine-producing cells. [70] For example, some works have shown a histamine role in the proliferation of human adrenocortical cells because it causes an inhibition of their proliferation, although no effect has been related with steroidogenesis. [71]

### 2.4.2. Muscle Endurance

Murine assays revealed an essential role of histamine in the muscle activity during physical exercise. Mice practicing prolonged walking revealed an increase of mHDC mRNA expression and mHDC activity in quadriceps femurs muscles.

Furthermore, mice supplied with antagonists of the $H_1$ or $H_2$ receptors or an irreversible mHDC inhibitor displayed less prolonged walking endurance than control group. [72]

#### 2.4.3. Hair growth

The hair-inducing ability of newborn mice dermal cells was studied and results showed a disappearance in the first few days of life, and also an mHDC expression decrease during that period. mHDC downregulation in newborn mouse dermal cells decreased their hair follicle-inducing ability, indicating a key role of mHDC in the maintenance of hair-inducing ability during the first few days of life in mice. [73]

#### 2.4.4. Granule Maturation

As mentioned before, mHDC is essential for the stored histamine production, mainly the 54 kDa form. Additionally, mHDC was also described as having a key role in the histamine-containing granules maturation [74], once mHDC$^{-/-}$ mutant mice showed an aberrant morphology with severely decreased granule contents. [75]

### 2.5. Drug target

The role of histamine in diverse physiological events is, unfortunately, associated with many diseases. This scenario turns mHDC as an important drug target to treat those pathologies. Consequently, the correct understanding of the catalytic mechanism of mHDC is crucial to unveil specific features about the enzyme, which could be used in the design of new drugs.

#### 2.5.1. Diseases

*Atopic Dermatitis*

Histamine $H_1$ receptor is the best-characterized receptor, and inverse agonists commonly referred to as antihistamines have been available for clinical use since 1942. There is a strong evidence for their use in the treatment of allergic rhinitis, allergic conjunctivitis and urticaria. In contrast, the use of antihistamines for the treatment of patients with atopic dermatitis is not currently recommended, based on their limited clinical efficacy and the risk of side-effects. [76-79] mHDC is expressed by epidermal keratinocytes, and its expression is significantly elevated in the skin of patients with atopic dermatitis. [79] mHDC expression by keratinocytes is also increased in response

to T-helper 2 cytokines and exogenous factors that play a role in triggering and exacerbating atopic dermatitis, including house dust mite allergens and bacterial lipopolysaccharide. Keratinocytes also have the ability to respond to histamine, expressing $H_1$, $H_2$ and $H_4$ receptors, which together orchestrate changes in the expression of over 4000 genes, including those involved in inflammation, proliferation and differentiation. [80]

*Chronic Allergic Contact Dermatitis*

Histamine can also facilitate the creation of eczematous lesions in the chronic allergic contact dermatitis murine model because of its ability to reduce the regulatory T cell population in the lesion. Consequently, the tissue lacks the cell population that is responsible for the attenuation and control of the inflammatory response. [81] mHDC could, therefore, be an interesting pharmacological target to treat chronic allergic contact dermatitis.

*Allergies*

Basal expression of mHDC is higher in neutrophils from patients with allergies than from healthy donors. This evidence shows that neutrophils could be a relevant source of histamine in allergic patients, causing this pathology. [5]

*Allergic Rhinitis*

Certain polymorphisms in the mHDC gene are associated with various histamine-related diseases and were presumed to affect the enzyme activity. Recently, it has been described a polymorphic difference in the mHDC gene, where the homozygous substitution of a glutamic acid at the position 644 by an aspartate increases the risk of developing allergic rhinitis. [82]

*Gastric Ulcer*

Histamine production and release has been implicated in the pathogenesis of gastric ulcers produced by restraint and pylorus ligation. [83, 84] The important role of histamine in the gastric ulcer pathology was essential for the research about mHDC as well as allergies and mastocytosis. [85]

Gastrin stimulation of ECL cells led to an elevated mHDC activity and increased secretion of histamine the major gastric acid secretagogue. The gastrin-stimulated

increase in mHDC activity has been shown to result in part from an increase in mHDC gene transcription in the gastric corpus and isolated ECL cells. [86, 87]

Bile acid secretion also increases the expression of mHDC as well as FXR (farnesoid X receptor), CDX1 (caudal-type homeodomain transcription factor) and SHP (small heterodimer partner). SHP was also found to be a primary regulatory transcription factor for bile acid-induced mHDC expression. [88]

*Diabetes*

A recent study has shown, recurring to a murine model of type-1 diabetes, that HDC knockout (KO) can decrease the incidence of diabetes when compared with wild-type mice. The HDC KO maintains the proportion of regulatory T cells and myeloid-derived suppressive cells in contrast with the increased levels of immature macrophages observed in KO mice. This study supports the idea that histamine is involved in the pathogenesis of diabetes. [89]

*Age at Natural Menopause*

Some HDC gene polymorphisms are likewise significantly associated with age at natural menopause (AANM) in Caucasian women. However, no pattern or explanation was described to justify this phenomenon. [90]

*Chronic Heart Failure*

Histamine release has also been described as a crucial role in the chronic heart failure (CHF), putting together the idea that the production of histamine inhibitors could be used to treat arrhythmias, sudden cardiac death, myocardial ischemia, and congestive heart failure. [91, 92]

The HDC role in the CHF pathology was also depicted, because a specific HDC allele (rs17740607) is associated with a decreased risk for CHF, as a result of the HDC activity decreasing and lower endogenous histamine levels in cells. [93]

*Cancer*

Histamine biosynthesis has been detected in a number of different tumors, including human small cell lung carcinomas [94], pancreatic tumors [95], colonic cancer cells [96], and human melanoma cells [97]. Moreover, in breast cancer, whereas histamine *N*-methyltransferase and histamine $H_3$ receptor appeared to be irrelevant in

the context of this cancer, a certain HDC allele (T allele of rs7164386) showed a significantly decreased of the incidence risk. On the other hand, rs7164386-rs7182203 alleles were more frequent among breast cancer patients. Polymorphisms of HDC gene could be significantly associated with breast cancer. [98]

A study about human colorectal cancer revealed that mHDC is highly expressed in the tumor microenvironment. Human colorectal cancer patients with lymph node invasion had a higher expression of mHDC than control. These results suggest that histamine expressed in microvessels could be an additional cellular source and involved in the cancer invasion through promoting angiogenesis in human colorectal cancer. [99]

An autocrine loop, consisting of enhanced histamine synthesis by mHDC, sustains cholangiocarcinoma growth is described, indicating HDC inhibitors as a promising treatment for cholangiocarcinoma patients. [100]

### 2.5.2. HDC inhibitors

Due to the pathologies described above, there has been a great interest in the development of HDC inhibitors to preclude undesired effects of histamine under certain circumstances. In clinical practice, there are only three types of approaches to inhibit the histamine effects, and none of them involved the inhibition of mHDC. Instead, the inhibitors prevent the degranulation of basophils and mast cells, blocking histamine release; behave as antagonists or inverse agonists of the histamine receptors; or act as antagonists that promote opposite effects without interfering with the normal histamine pathways. Nonetheless of these strategies, a better choice would be the usage of compounds that would inhibit mHDC directly, decreasing the effect of histamine more efficiently. In fact, several inhibitors have been studied, but none of them has turned out to be suitable to be used in clinical practice.

The first studies regarding the direct inhibition of mHDC *in vitro* were devoted to norepinephrine (**Fig. 17**) and several structurally related compounds. These compounds are reversible inhibitors of HDC that compete with the natural substrate to form similar EA. All of those compounds share a structural pattern of an *m*-hydroxyl-phenethylamine (**Fig. 17**) with a terminal primary amine. This amine group can form an EA with the PLP cofactor, but they could easily unbind the PLP again leading to a reversible competitive type of inhibition. When a substitution of either a methyl or a carboxyl group on the α-carbon is made, the inhibitory potency slightly changes. However, the simultaneous substitution leads to a complete inactivation of the mHDC. This irreversible inhibition is caused by the decarboxylation of the inhibitor and the consequent greater stabilization

of the resulting QI. Authors suggested that the inhibitors condensation with PLP was crucial to an efficient inactivation of the enzyme. [101]

4-bromo-3-hydroxybenzyloxyamine dihydrogen phosphate (NSD-1055) (**Fig. 17**) is also a potent inhibitor of mHDC and it can reduce histamine excretion in the urine after oral administration. Once again, NSD-1055 can form an EA with the cofactor and prevent the natural substrate bind. [102] Following studies assessed the potency of aliphatic and benzyl aminooxy amines (NSD-1055 derivatives) into the inhibition of mHDC and all of them were tested and showed at most 100-fold less potency compared with NSD-1055. [16] Although less potent, these inhibitors helped new searches to find new inhibitors capable to reduce mHDC activity in a selective manner, because NSD-1055, for example, also exhibited the capability to inhibit dopamine-β-hydroxylase. [103]

Fewer years late rhodanine (**Fig. 17**) is described as a potent inhibitor of gastric rat HDC *in vitro* and the mechanism of inhibition involves the reaction of this compound with the PLP cofactor and the condensation reaction between 5-methylene carbon of rhodanine and enzyme-bond PLP, once rhodanine can inhibit only the holoenzyme and not the apoenzyme. [104] These data also corroborate the idea that efficient inhibitors should bind the PLP cofactor.

An inhibitor must be not only potent but also specific. However, the inhibitors described above are not so specific as desired. In 1978, the 4-imidazolyl-3-amino-2-butanone (McN-A-1293) (**Fig. 17**) was produced with the intent to be the first specific inhibitor of mHDC. [105] Indeed, McN-A-1293 can inhibit mHDC *in vitro* but not AADC, and it is also a weak inhibitor of DAO and HNMT, which are enzymes that degrade histamine when it is released. [106] This high specificity is due to a great similarity with the natural substrate. Nevertheless, McN-A-1293 is a weaker inhibitor of mHDC and researchers tried to synthesize other histidine derivatives to interfere more efficiently with mHDC activity. [107] These compounds contain different substituents in the α-carbon or/and adding different ester groups. Consequently, a good inhibition was achieved with the methyl ester of *L*-histidine (HME) (**Fig. 17**) with an inhibition of 50% at $4\times10^{-6}$ M concentration with a $K_i$ = $1{,}8\times10^{-6}$. Analogous esters with ethyl and other alkyl groups were tried, but it caused a significant decrease of the inhibitor efficiency. Also, no substitution in the α-carbon evidenced a powerful modification to increase the inhibition of mHDC. Moreover, amide derivatives did not reveal a capability to inhibit mHDC. [107] At this moment, HME is the most specific and efficient inhibitor of the mHDC activity, and only requires an esterification of the carboxylate group with a methyl group.

Later on, a new work showed the ability of some histidine derivative peptides to inhibit mHDC with a potency about 100-fold greater than the previous HME. It was

observed that histidine-phenylalanine dipeptide (**Fig. 17**) was the most effective peptide inhibitor, closely followed by histidine-tyrosine dipeptide (**Fig. 17**). These dipeptides can form the EA and block the binding pocket for the natural substrate. Furthermore, both peptides are unable to inhibit AADC as well as HME, presenting a similar specificity comparatively with HME. [108] This data indicates that the increase of hydrophobicity of an inhibitor increases their potency toward mHDC. [109]

3-Methoxy-5,7,3',4'-tetrahydroxyflavan (ME) (**Fig. 17**) was reported as gastric anti-ulcer drugs because it is also able to inhibit specifically the mHDC in rats and guinea-pigs. Results showed a real capability of ME to inhibit mHDC and reduce the amount of histamine produced in the stomach. [85] The mechanism of inhibition is not known, but it is believed that the high hydrophobicity places the inhibitor at the binding pocket, hindering by this way the substrate binding.

The most powerful mHDC inhibitor that is known is (S)-α-fluoromethyl-histidine (FMH) (**Fig. 17**). The inhibitory mechanism of this compound was extensively studied because FMH can irreversibly inactivate mHDC [110, 111] and the bacteria form from *Morganella morganii* [112, 113] but had no effect on pyruvyl-dependent HDC such as that from *Lactobacillus 30a*. [110, 114] Many groups have shown that mHDC inhibition by FMH is accompanied by the decarboxylation of the inhibitor and consequent fixation to the enzyme. [111, 115] The fixation occurs by β-elimination of fluoride. Unfortunately, biological assays showed several undesired side effects when FMH is used to inhibit mHDC, such as sympathetic facilitations, arterial hypertension and tachycardia. [116]

The development of FMH, and HME, as new mHDC inhibitors have been very useful for basic research on mHDC enzymology, histamine physiology, and molecular biology, but they are not suitable drugs for human therapy. One of the problems is that these two analogs are also able to inhibit the gram-negative bacterial enzyme giving rise to important secondary effects because of their presence in the intestinal track. [67]

Another important inhibitor of mHDC is an abundant catechin in green tea, i.e., epigallocatechin-3-gallate (EGCG). EGCG has been shown to target histamine-producing cells and to promote anti-inflammatory, antitumor and anti-angiogenic effects. Some experimental works have demonstrated that EGCG has a direct inhibitory effect on both mHDC and AADC. [117-125] EGCG is able to inhibit mHDC by partially occlusion of substrate entrance and establishes several hydrogen bonds with the highly conserved residue of the active site, namely His-194A and several other residues in a functional loop of the opposite monomer (Tyr-334B-conatining loop). As the active site of AADC is more exposed to the solvent than the active site of histidine decarboxylase, EGCG can make more electrostatic interactions in the binding pocket with His-192 (the equivalent

of mHDC His-194A) and Leu-333B in the homologous functional loop and even with the cofactor. These first results confirmed that EGCG interacts with the binding pocket, and it does not change the dimeric structure of the enzyme. This blockage was confirmed by a molecular docking, where *L*-histidine, HME, and FMH were docked into an mHDC previously inhibited with EGCG. The results revealed that none of them were able to bind to the mHDC binding site. In conclusion, the mechanism of inhibition of EGCG is based on alterations of the chemical environment at the active site and physical occlusion of the substrate entrance. So, no distortion of the quaternary structure or dimerization surface is induced by EGCG and other polyphenols. [126] The effect of EGCG was tested in an animal (guinea pig) allergic conjunctivitis model in combination with cetirizine ($H_1$ receptor antagonist) The combination of catechin and cetirizine showed a statistically significant effectiveness in the management of allergic conjunctivitis when compared with the standard disease control and catechin and cetirizine individual groups. [127]

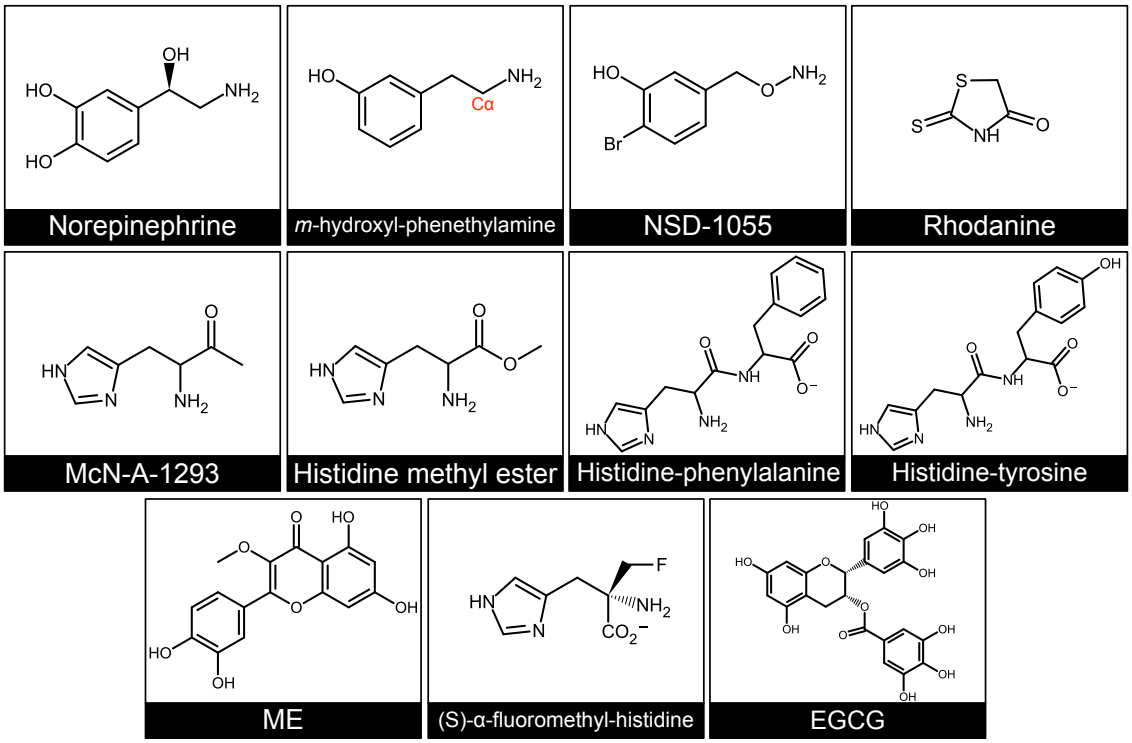



**Fig. 17 – Overview of the structure of all inhibitors described above.**

Other studies pointed out the dimerization surface as the most promising area for a specific inhibition of mHDC because the bacterial enzyme seems to differ mainly in its quaternary structure from the mammal ones. [128] Especially, the flexible loop, comprising residues 331–349 in the rat enzyme, has no homologous counterpart in the bacterial enzyme, and it is close to the entry of the catalytic pocket of the enzyme. [129] Therefore, a computational alanine-scanning mutagenesis (compASM) study could be

an interesting approach aiming the interface area between the two subunits. The identification of relevant amino acid residues for the dimerization could be used as drug targets intending the mHDC inhibition.

# B. Theoretical Background

*He who loves practice without theory is like*
*the sailor who boards ship without a rudder and a compass and*
*never knows where he may cast.*
**Leonardo da Vinci**

This page was intentionally left blank.

Theoretical and computational means are nowadays often used to study several aspects of biological systems. This, in part, results from the methodological advances that allow simulating with high accuracy aspects of these systems, but also due to the technological advances that permit the use of these methods in big and complex systems, without a significant computational cost.

Experimental data reveals important information about the systems however, in case of catalysis, they cannot assess data about the cleavage and formation of chemical bonds. Nevertheless, they provide very relevant information about complex structures and intermediates, which give indispensable clues about the mechanism. Computational studies can then be used to unravel the mechanism that leads to the formation of those reaction intermediates and provide the complete “movie” of the full catalytic process.

Currently, there are four theoretical levels of theory that are commonly used to study enzymatic mechanism, and their usage has a direct impact on the accuracy of the calculation, the number of atoms that can be used and computational cost. They are normally classified as *ab initio*; Density-Functional Theory (DFT), semi-empirical; and Molecular Mechanics (MM).

The *ab initio* methods are the most reliable ones to describe a molecular system. Among the *ab initio* methods, Hartree-Fock (HF) and Pos-Hartree-Fock (Pos-HF) are the most used ones. The Pos-HF methods are the most accurate ones but require high computational effort, and, therefore, can only be used with a few dozens of atoms. In this work, we only will concern about the HF methods because, although they are not so accurate in relation to the Pos-HF methods, they allow the use of a wider range of atoms in the calculations. Density-Functional Theory (DFT) approach allowed a description of biological and chemical systems almost so accurate as the *ab initio* methods but significantly less computationally demanding. These methods allow therefore to increase the complexity of the model systems that are studied. A less demanding approach can be achieved through semi-empirical methods. These methods are based on HF formalism, but introduce several approximations and parameters that also turn them considerably less accurate. Finally, the MM methods are based on classical physics to predict atomic interactions and movements. However, these methods do not include electrons in their description and consider atoms as non-deformable spheres with charges and linked each other by springs. For this reason, they cannot be used to simulate chemical reactions and phenomena that involve electronic events. The huge advantage of these methods is their ability to be used in systems with many atoms (thousands) with an inexpensive computational cost. All these approaches will be discussed in the next sections, particularly the methods used in this work.

More recently, we have been witnessing the development of hybrid methods which are the conjugation of two different types of above theoretical methods. Hybrid methods, as QM/QM and QM/MM, emerged from the necessity of study large systems, as biological complexes, with a compromise between accuracy and computational cost. Later on, these methodologies will also be described in more detail.

# 1. Molecular Mechanics

Molecular mechanics (MM), by definition, is used to refer the use of simple potential-energy functions as the harmonic oscillator and Coulombic potentials in molecular systems. [130]

In MM, the smallest particle of the system is the atom, which is treated as a charged sphere linked to other atoms by springs representing bonds. All interactions are then calculated using simple mathematic approaches. For example, bonded interactions are calculated using the Hook's law, whereas the non-bonded interactions, such as, the electrostatic and van der Waals interactions are treated using Coulombic and Lennard-Jones potentials, respectively. Then, the system energy can easily be calculated based on bond stretching, angle bending, dihedral torsion, Coulombic and van der Waals interactions. [130]

We are unable to use MM to study chemical reactions because it involves the cleavage and formation of bonds, which requires an electronic description. Therefore, MM is a powerful tool to explore the conformational space of a system, for example when it is used in simple conformational searches of in more complex Molecular Dynamics (MD) or Monte Carlo simulations. [131]

## 1.1. Force Fields

As we see above, MM does not use wave function into the energy calculation. Instead, the energy is calculated based on a simple mathematic equation which requires some parameters obtained by experimental works or *ab initio* calculations. The MM force fields allow "to approximate" quantum mechanical energy surface, using less computational resources. [132, 133]

Generally, force fields use a class I additive potential energy function to achieve the system energy. In this type of functions, all terms for the bond interactions are added to all terms for the non-bonded interactions (**Equation 1**). [132]

$$V_{system} = V_{bonded} + V_{non-bonded}$$

**Equation 1 - Class I additive potential energy function**

Different force fields have been developed and optimized for particular types of systems. AMBER (Assisted Model Building with Energy Refinement) [134] is one of the most used software that includes ff99SB and GAFF (General AMBER force field) force

fields [135] which is used to protein or nucleic acid-containing systems. Nevertheless, there are several other force fields such as CHARMM (Chemistry at Harvard Macromolecular Mechanics) [136], OPLS (Optimized Potentials for Liquid Simulations) [137] or GROMOS (Groningen Molecular Simulation package) [138].

In this work the AMBER ff99SB and GAFF force fields [135] were used so that it will be described below.

### 1.1.1. Bonded interactions

Considering the AMBER force fields [135], the bond interactions are divided into three categories: the bond stretching terms, the angle bending terms, and the dihedral terms (**Equation 2**).

$$V_{bonded} = \sum_{bonds} V_{bond} + \sum_{angles} V_{angle} + \sum_{dihedrals} V_{dihedral}$$

**Equation 2 – Energy function for the bond interactions.**

The bond stretching contributions to the potential energy are approximated by a harmonic oscillator as a function of the bond length ($l$) (**Equation 3**). [133, 135]

$$V_{bond} = \frac{1}{2} k_b (l - l_0)^2$$

$k_b$ is the force constant; $l$ is the bond length; $l_0$ is the bond length at the equilibrium.

**Equation 3 – Bond-stretching potential energy ($V_{bond}$) function.**

A similar harmonic oscillator approximation is also used to describe the angle bending potential energy as a function of the angle between three adjacent atoms ($\theta$) (**Equation 4**). [133, 135]

$$V_{angle} = \frac{1}{2} k_a (\theta - \theta_0)^2$$

$k_a$ is the force constant; $\theta$ is the angle between three adjacent atoms; $\theta_0$ is the angle between three adjacent atoms at the equilibrium.

**Equation 4 – Angle-bending potential energy ($V_{angle}$) function**.

For these two potential energy calculations, only two parameters are needed for each one, the force constant ($k_b$ or $k_a$) and the bond length ($l_0$) or angle value ($\theta_0$).

The dihedral angle term, also called torsion angle, corresponds to the energy associated with the rotation of atoms separated by three bonds. In other words,

considering atoms A, B, C and D bonded each other in this order, the dihedral angle is the angle between planes ABC and BCD. The potential energy associated with the dihedral angle is given as a cosine function by the **Equation 5**. [133, 135]

$$V_{dihedral} = \frac{1}{2} k_d (1 + \cos(n\omega + \gamma))$$

$k_d$ defines the energy barrier for the dihedral rotation; $\omega$ is the dihedral angle and could assume values between 0° and 360°. $\gamma$ corresponds to the dihedral angle value for which the energy is minimum; $n$ defines the number of function's minimums.

**Equation 5 – Dihedral angle potential energy ($V_{dihedral}$) function.**

### 1.1.2. Non-bonded interactions

The non-bonded interactions might be split into electrostatic and van der Waals interactions (**Equation 6**). [133, 135]

$$V_{non\ bonded} = V_{elect.} + V_{van\ der\ Waals}$$

**Equation 6 – Energy function for the non-bonded interactions.**

The electrostatic interaction describes non-bonded interaction as a result of charges or permanent dipoles. The Coulombic law is used to the analytic calculation considering atoms as charged points separated by their interatomic distance (**Equation 7**). [133, 135]

$$V_{elect.}(i,j) = \frac{1}{4\pi\varepsilon} \frac{q_i q_j}{r_{i,j}}$$

$q_i$ and $q_j$ are the atomic charges of the atoms i and j, respectively; $r_{i,j}$ is the distance between atoms i and j; $\varepsilon$ is the dielectric constant.

**Equation 7 – Electrostatic potential energy ($V_{elect.}(i,j)$) function between atoms i and j.**

The van der Waals interaction describes repulsive and attractive non-bonded interactions, which are not described by the electrostatic term, such as nuclear repulsion (Pauli's repulsion force), instantaneous and induced dipoles attraction (London forces). In the AMBER force fields, the van der Waals potential energy is calculated using a 12-6 Lennard-Jones potential as a function of the distance between two atoms (**Equation 8**). [133, 135]

$$V_{van\ der\ Waals}(i,j) = \varepsilon_{i,j}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)$$

$\varepsilon_{i,j}$ is the function minimum; $r_{i,j}$ is the distance between atoms i and j; $\varepsilon$ is the dielectric constant; $\sigma_{ij}$ is the interatomic distance for which the energy is zero.

**Equation 8 – Van der Waals potential energy ($V_{van\ der\ Waals}(i,j)$) function between atoms i and j.**

## 1.2. Molecular Dynamics

Molecular Dynamics (MD) is a common tool to assess relevant information about the conformational space of a chemical system. It can also provide important thermodynamic and structural flexibility information.

Since the first MD simulations made in 50's [139], they have been used for several different purposes, such as protein stability, molecular docking, conformation rearrangement studies, protein folding, transport through the cellular membrane, among others.

A critical point when we want to choose a duration time for a MD simulation is the commitment between the computational cost and the movements that we want to sample. For example, atomic fluctuations occur into the fento and picoseconds time scale; loops and terminal regions movements take between nano to microseconds to happen, and subunit movements and protein folding could exceed some microseconds. In this work, we performed some nanoseconds of MD simulation to assess information about protein flexibility and atomic fluctuations.

A MD simulation predicts the atom movement based on integration of Newton's equation from which results data about velocity and position for each atom. At each time point, a differential equation derivative from the Newton's Second Law (**Equation 9**) is defined for each system's particle. [133]

$$\frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i}$$

**Equation 9 – Differential equation that describes the motion of a particle with mass $m_i$, along one coordinate $x_i$ with a force $F_{x_i}$.**

Due to the difficulties behind the analytical solution for that equation, a finite integration step ($\Delta t$) is used to numerically calculate the position (**Equation 10**) and the velocity (**Equation 11**) for each atom in the system.

$$\overrightarrow{x_i}(t_0 + \Delta t) = \overrightarrow{x_i}(t_0) + \overrightarrow{v_i}(t_0)\times\Delta t + \frac{1}{2}\,\overrightarrow{a_i}(t_0)\times\Delta t^2$$

$\overrightarrow{x_i}(t_0)$ is the atomic position in the time $t_0$; $\overrightarrow{v_i}(t_0)$ is the atomic velocity in the time $t_0$; and $\overrightarrow{a_i}(t_0)$ is the atomic acceleration in the time $t_0$.

**Equation 10 – Atomic position calculation in the $t_0 + \Delta t$ time point using the finite integration step $\Delta t$.**

$$\overrightarrow{v_i}(t_0 + \Delta t) = \overrightarrow{v_i}(t_0) + \overrightarrow{a_i}(t_0)\times\Delta t$$

$\overrightarrow{v_i}(t_0)$ is the atomic velocity in the time $t_0$; and $\overrightarrow{a_i}(t_0)$ is the atomic acceleration in the time $t_0$.

**Equation 11 – Atomic velocity calculation in the $t_0 + \Delta t$ time point using the finite integration step $\Delta t$.**

As the above equations suggest, three different variables must be known at the beginning: position, velocity, and acceleration. Atomic positions are obtained, for example, by the X-ray crystallography or NMR. The velocity distribution of all atoms in the system is done recurring to a random assignment from a Maxwell-Boltzmann distribution at the system temperature. The acceleration is gotten by the ratio of the force and the atomic mass, once the force could be achieved by the potential energy derivative in order to the particle coordinates. [133]

### 1.2.1. Integration step

As mentioned above, we need to set an integration step that defines the time interval between position and velocity calculations for all atoms. Using the smallest integration step is the most desirable situation because it minimizes the errors and becomes closely to the analytical solution. However, the smaller is the integration step, the highest will be the computational cost.

To overcome this situation, it is a good practice use a ten times smaller integration time comparing to the fastest movement in the system. In our system, as well as other biological complexes, the fastest event is bond stretching of the hydrogen bonds and occurs with time lapses of approximately 10 fs. Then, an integration time of 1 fs is enough and was used in the present work.

### 1.2.2. Ensembles

The statistical mechanics is a bridge between the molecular events and the macroscopic system properties. A macroscopic system is characterized by α+2 state functions, where α is the number of components. A state function is, for example, the temperature, volume, pressure, number of particles, among others. [133]

A macrostate could be defined as an ensemble of microstates under the same thermodynamic conditions. The most common types of ensembles are the canonic ensemble (number of particles (N), volume (V) and temperature (T) constants); isothermal-isobaric (N, pressure (P) and T constants); microcanonic (N, V and total energy (E) constants); and the grand canonic (V, T and chemical potential (μ) constants). [133]

In this project, the canonic (NVT) and the isothermal-isobaric (NPT) ensembles were used in the MD simulations for the equilibration and production moments, respectively.

### 1.2.3. Periodic boundary conditions

Usually, the system dimensions used in MD simulations are small because of the computational cost inherent. So, the studied system is composed of the protein inside a water box, normally modeled as a cubic box. The small models lead to a problem related to surface effects. Water molecules near the box surface are under different conditions comparing with water molecules inside the box. These different environments lead to additional errors that could be solved ignoring those molecules. However, those molecules could represent approximately 40% of the total number of molecules in the system, invalidating the ignoring approximation.

A way to overcome that is recurring to periodic boundary conditions, so the system is only a central spot in an infinite cubic network consisted by their cell replicas. The motion equations are only calculated for the central unit, and the movements are reproduced into all other units.

Evidently this is not an ideal solution because units are following the same oriented movements, but it is a good approximation to fix this pitfall. Nevertheless, we have to pay attention to prevent abnormal results, once the solvation box should be big enough to prevent protein atoms interact with other protein atoms from the replica units. For that, in this work, the solvation box was placed in a way that allows a minimum distance of 12.0 Å between the protein atoms and the unit surface, since we used a non-bonded interactions cutoff of 10.0 Å. The non-bonded interactions are the property that imposes a minimum size for the unit because they could be calculated for a non-finite number of atom interaction due to the infinite number of units.

### 1.2.4. Non-bonded interactions cut-off

The non-bonded interactions could be split into short and long range, where the first ones are habitually described by Lennard-Jones potentials. In this case, the van der Waals potential energy is calculated recurring to a 12-6 Lennard-Jones potential where the highest term is $r^{-6}$ dependent, so not far away atoms have already a little contribution for the potential energy. However, the Coulombic interactions have a $r^{-1}$ interatomic distance dependency, so more atoms have a non-neglected influence into the total potential energy.

The non-bonded calculation for all atoms from all replica units is not possible because there are an infinite number of interactions. Subsequently, an approximation has to be done to reduce the number of interactions calculated, and a truncation radius is defined. In this work, we use a cut-off radius of 10.0 Å, so all non-bonded interactions between each atom and all atoms that do not distance more than 10.0 Å are calculated. This approximation is not pejorative for the van der Waals interactions because after 10.0 Å they are practically insignificant. Nevertheless, it is desirable the correction factor adding to adjust this neglected interactions, once all van der Waals interactions are attractive, and a sum of thousands of those interactions could have a significant weight in the total energy calculation.

The Coulombic interactions are an important term indeed for the far interactions, but the attractive interactions with farthest atoms could easily cancel out by the repulsive ones resulting in a weak contribution for the energy calculation.

In conclusion, it is desirable the use of a high cut-off radius, but we should choose a radius that follows a good commitment between computation cost and accuracy.

# 2. Quantum Mechanics

When only the macroscopic world was known, the classical physics based on thermodynamics, electromagnetism and Newton's movement laws were able to correctly describe and explain those phenomena. Nevertheless, when the atomic and subatomic studies started happening, the classical physics equations could not be applied to those systems. Thereat, several scientists as Einstein, De Broglie, Planck, Bohr, Schrödinger, among others postulated theories about Quantum Mechanics (QM), which is able to describe the majority of all particle-containing system.

In the beginning, QM calculations and theories could only be applied to small systems as the hydrogen atom, particle in the box, harmonic oscillator, particle in the ring and particle in the sphere. [133] Afterward, the computational developments together with some mathematical approximations turn QM able to be set in more complex molecular systems.

The Quantum Theory born when Max Planck published work, in 1901, describing the concept of a quantum of matter and electricity, leading to a proportional relation between the energy and the radiation frequency (**Equation 12**). [140]

$$E = hv$$

$E$ is the energy of a photon with a certain frequency $v$. $h$ is the Max Plank's constant and it has the value $6.626069\times10^{-34}$ J s. [141]

**Equation 12 – Planck-Einstein relation.**

This concept allowed Einstein to explain the photoelectric effect [142] (discovered by Hertz [143]), considering an electromagnetic wave a set of particles. The interpretation of the light as a set of particles was confirmed later on by Compton in 1923. [144] Almost at the same time, Bohr presented his theory about the atomic structure, showing that electrons are also quantized, and they could only occupy certain energy levels in the atom. Consequently, the electron movement within the atom must be accomplished by emission or absorption of energy. [145] In 1924, Louis de Broglie confirmed the reciprocity of the wave-particle duality of the matter, affirming that any body has a wave associated with it, and it depends on its momentum (**Equation 13**).

$$\lambda = \frac{h}{p}$$

$\lambda$ is the wavelength of a certain body with a momentum $p$.

**Equation 13 – de Broglie wavelength.**

In 1927, Heisenberg introduced the notions of the "new" QM as the uncertainty principle [146], which criticizes the classical limitations of simple observation. The uncertainty principle indicates an impossibility to know accurately, for example, both the moment and the position of a certain particle at the same time, according to the **Equation 14**.

$$\sigma_x \sigma_p \geq \frac{\hbar}{2}$$

$\sigma_x$ and $\sigma_p$ are the position and momentum associated errors. $\hbar$ is the reduced Plancks' constant ($\hbar = h/2\pi$).

**Equation 14 – Uncertainty principle.**

Based on these works and others by Dirac, Fermi, Born, Davisson, Germer, Debye and Maxwell, for example, QM acquired the actual five fundamental postulates from which the first one can define the system mathematically: [147]

"To every state of a physical system, there is a function Ψ ascribed to and defining the state."

It means that each physical system, observed by a quantum point of view, is defined by a wave function depending on the spatial coordinates (x, y and z) and the time (t) which can determine all the features that can be known about the system. [147, 148] It is important to refer that the wave function represents the probability amplitude. In other words, if we consider a hydrogen atom from which the proton is the center of a reference system, the wave function gives the probability to find the electron as a function of the electron position. [147]

The second postulate defines how we could get information about the system, recurring to operators, where a certain operator could be used to assess certain system property (**Equation 15**). [133, 147, 148]

$$\hat{A}\Psi = e\Psi$$

Ψ is the wave function; $\hat{A}$ is the operator for a certain system property; and $e$ is the eigenvalue for that property.

**Equation 15 – Second Quantum Mechanics Postulate.**

The **Equation 15** is correlated to the third postulate that defines the eigenvalue as the only information that can be measured from the system when a certain operator is used. It means that each single measurement corresponds to a quantum operator ($\hat{A}$)

and only the eigenvalues of the operator are measured satisfying **Equation 15**. [147, 148]

## 2.1. Schrödinger Equation

The Schrödinger equation is a derivative equation from the **Equation 15** where the operator is the Hamiltonian, and the given eigenvalue is the total system energy. Whether the time-dependent Schrödinger equations (**Equation 16**) could be analytically solved, then all system time evolution could be predicted. [133, 147, 148]

$$\widehat{H}\Psi(r,t) = E\Psi(r,t)$$

$\Psi(r,t)$ is the wave function of the particle with respect to the position ($r$) and time ($t$); $\mu$ is inertial mass; $\widehat{H}$ is the Hamiltonian (quantum operator); $E$ is the energy of the particle.

**Equation 16 – Time-dependent Schrödinger equation.**

In **Equation 16**, the second member could be defined as the product of imaginary unit, reduced Planck's constant and the partial derivative with respect to time of the wave function (**Equation 17**). [133, 147, 148]

$$\widehat{H}\Psi(r,t) = i\hbar\frac{\delta}{\delta t}\Psi(r,t)$$

$\Psi(r,t)$ is the wave function with respect to the position ($r$) and time ($t$); $i$ is the imaginary unit; $\hbar$ is the Plank constant divided by $2\pi$.

**Equation 17 – Time-dependent Schrödinger equation.**

The Hamiltonian ($\widehat{H}$) could also be described as the sum of a kinetic operator ($\widehat{T}$) and a potential operator ($\widehat{V}$) (**Equation 18**).

$$\widehat{H} = \widehat{T} + \widehat{V}$$

**Equation 18 – Hamiltonian is a function of kinetic and potential operators.**

The potential operator $\widehat{V}$ could be described as the potential energy ($V(r,t)$). For example, if we consider a single electron under the potential field of an atomic nucleus with Z protons, the time-independent potential energy could be defined by **Equation 19**.

$$V(r) = \frac{Ze^2}{4\pi\varepsilon_0 r}$$

$Z$ is the number of protons in the atomic nucleus; $e$ is the electron charge; $\varepsilon_0$ is the electric permittivity of vacuum; $r$ is the distance between the electron and the atomic nucleus.

**Equation 19 – Potential energy calculation for an electron under the potential field of an atomic nucleus.**

On the other hand, the kinetic operator is calculated based on the Laplacian operator (a divergence of the gradient of a $x$, $y$ and $z$ function — $\left(\frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2} + \frac{\delta^2}{\delta z^2}\right)$) and inertial particle mass.

Then, we could write the Schrödinger equation in an extended way (**Equation 20**).

$$\left[\frac{-\hbar^2}{2\mu}\left(\frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2} + \frac{\delta^2}{\delta z^2}\right) + V(r,t)\right]\Psi(r,t) = i\hbar\frac{\delta}{\delta t}\Psi(r,t)$$

$\Psi(r,t)$ is the wave function with respect to the position ($r$) and time ($t$); $\mu$ is inertial mass; $x$, $y$ and $z$ are the Cartesian coordinates; $V(r,t)$ is the potential energy with respect to the position and time; $i$ is the imaginary unit; $\hbar$ is the Plank constant divided by $2\pi$.

**Equation 20 – Time-dependent Schrödinger equation.**

If the potential function was time-independent, we could write the Schrödinger equation as a function that depends only on the position of the particle (**Equation 21**). [133]

$$\widehat{H}\Psi(r) = E\Psi(r)$$

**Equation 21 – Time-independent Schrödinger equation.**

The wave function does not decode into a physical property, but its square represents the probability density at a certain time. If it is multiplied by a spatial volume, the result is the probability to find that particle in that position. [149]

## 2.2. Hamiltonians

Mathematically operators are used to get information about system properties from the wave function. The operator application to a eigenfunction results into a wave function multiplied by the correspondent eigenvalue. In this way, the Hamiltonian ($\widehat{H}$) is the mathematical operator for the wave function that gives the total system energy. [133]

As described in the Equation 18, the Hamiltonian is the result of the sum of the kinetic and the potential energy operators, which can be expanded according to **Equation 22**. [133]

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn}$$

$\hat{T}_n$ and $\hat{T}_e$ are the kinetic operators for nuclei and electrons, respectively. $\hat{V}_{ne}$, $\hat{V}_{ee}$ and $\hat{V}_{nn}$ are the potential operators for the nucleus-electron, electron-electron and nucleus-nucleus interactions, respectively.

**Equation 22 – Hamiltonian is a function of kinetic and potential operators.**

Considering the two particles system electron-proton, the kinetic energy operator ($\hat{T}_n$ and $\hat{T}_e$) could be calculated according to **Equation 23**.

$$\hat{T}_e = \left[\frac{-\hbar^2}{2\mu_{electron}}\left(\frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2} + \frac{\delta^2}{\delta z^2}\right)\right]_{electron} ; \hat{T}_n = \left[\frac{-\hbar^2}{2\mu_{nucleus}}\left(\frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2} + \frac{\delta^2}{\delta z^2}\right)\right]_{nucleus}$$

**Equation 23 – The kinetic energy operator ($\hat{T}$) for an electron or a nucleus.**

Additionally, the potential operator is given by the attractive energy associated with the attraction between the nucleus and the electron. If we take in consideration more than one atom, this term is also affected by the repulsion interactions between different electrons and between different atomic nuclei (**Equation 24**). The potential operators are calculated using a Coulombic-like potential as the MM calculations. [133]

$$\hat{V}_{nn} = \sum_{A-B}^{N} \frac{Z_A Z_B}{r_{AB}}; \hat{V}_{ne} = \sum_{A}^{N}\sum_{i}^{M} \frac{Z_A Z_i}{r_{iA}}; \hat{V}_{ee} = \sum_{i-j}^{M} \frac{Z_i Z_j}{r_{ij}}$$

$A$ and $B$ are two different atomic nuclei; $i$ and $j$ are two different electrons; $N$ and $M$ are the number of nuclei and electrons, respectively; $Z$ is the charge of each particle; $r$ is the distance between two particles.

**Equation 24 – The potential operator ($\hat{V}$) for a multi-atomic system.**

The exact analytically solution for a Schrödinger equation within a poli-electronic system is mathematically impossible. Therefore, some approximations and considerations must be done to make it available to be applied in more complex systems.

## 2.3. Variational Principle

The variational principle is an approximation aiming a solution to the Schrödinger equation, and it uses an arbitrary wave function able to describe the electronic and nuclear coordinates appropriately and also being operated upon by the Hamiltonian. It is

mathematically possible to prove that the energy value obtained by the wave function is always greater or equal to the real fundamental state's energy (**Equation 25**). [149]

$$\frac{\int \Psi_a \hat{H} \Psi_a \, d\tau}{\int {\Psi_a}^2 \, d\tau} \geq E_0$$

**Equation 25 – Variational principle. $\Psi_a$ is an arbitrary wave function and $E_0$ is the real fundamental state's energy.**

The relation in **Equation 25** is very important because it allows the test of several arbitrary wave functions and reveals which ones have the lowest value. The lower is the value of the variational integral and closer will be this value to the real one. Consequently, the most accurate will be the wave function that will provide the better description of the closely "true" wave function. Taking this into account, we do not need to construct the wave function using a linear combination of orthonormal wave functions. Instead any random wave function could be built and tested. [149]

## 2.4. Born-Oppenheimer Approximation

Besides the variational principle consideration, the Hamiltonian calculation for more complex molecular systems leads to the requirement of some more mathematically approximations to make calculations feasible.

In 1927, the Born-Oppenheimer approximation was described as a manner to simplify the wave function calculations with a tremendous relevance for molecular systems. Born-Oppenheimer approximation postulates that the motion of the atomic nuclei and electrons can be separated. It means that the total wave function could be a result of the nuclei wave function product by the electrons wave function (**Equation 26**). [150]

$$\Psi_{Total} = \psi_{nuclear} \times \psi_{electronic}$$

**Equation 26 – General Born-Oppenheimer approximation equation.**

This is an acceptable approximation taking into account that the proton ($1.67 \times 10^{-27}$ kg) is 1836 heavier than the electron ($9.11 \times 10^{-31}$ kg) leading to an almost instantaneous electron rearrangement after nuclear movement. [133, 151]

Then, we could apply this approximation in the Hamiltonian calculation and split the total system energy into the nuclear and electronic energies (**Equation 27**).

$$\widehat{H}_{Total} = \widehat{H}_{nuclear} + \widehat{H}_{electronic}$$

**Equation 27 – The Born-Oppenheimer approximation in a Hamiltonian form.**

The electronic Hamiltonian is calculated with nuclei positions fixed, considering kinetic energy of each electron under the nuclei's force field ($\hat{T}_{elec.}$), the potential energy involved in the interelectronic repulsion ($\hat{V}_{elec.-elec.}$) and in the nucleus-electron attraction ($\hat{V}_{elec.-nucleus}$) (**Equation 28**).

$$\widehat{H}_{electronic} = \hat{T}_{elec.} + \hat{V}_{elec.-elec.} + \hat{V}_{elec.-nucleus}$$

**Equation 28 – Electronic Hamiltonian following the Born-Oppenheimer approximation.**

On the other hand, the nuclear Hamiltonian is calculated by the sum of the kinetic energy associated with the nuclear movements ($\hat{T}_{nucleus}$) and the potential energy related with the internuclear repulsion ($\hat{V}_{nucleus-nucleus}$) (**Equation 29**).

$$\widehat{H}_{nuclear} = \hat{T}_{nucleus} + \hat{V}_{nucleus-nucleus}$$

**Equation 29 – Nuclear Hamiltonian following the Born-Oppenheimer approximation**

### 2.4.1. Electronic Hamiltonian

In the study of catalytic mechanism in biological systems, the electronic Hamiltonian presents the biggest contribution to the total energy comparing to the nuclear contribution. So, the nuclei position is often fixed, and the energy is calculated based only on the electronic fraction. Several theories and methodologies were developed to calculate the electronic energy contribution in a more accurate and less expensive way possible.

#### *2.4.1.1. Wave function-based Theories*

Considering the wave function-based theories, we could distinguish three different theories, the Hartree-Fock (HF), the post-Hartree-Fock (pos-HF) and the semi-empirical theories. All of these theories were developed in order to calculate the electronic fraction of the Schrödinger equation, creating a Potential Energy Surface (PES) from which we could obtain, for example, equilibrium geometries and vibrational frequencies. The HF methodologies result in more exact solutions, but they are computational more expensive. The semi-empirical theories are derived from the HF models, but these neglect all mathematically integrals involving more than two nuclei in

the Fock matrix consideration. The success of the semi-empirical theories usage depends on the accurate substitution of the neglect integrals by parameters fitting experimental results. Semi-empirical methods are substantially more computationally efficient, but their applications are conditioned by the existing parameterization. Each of these methods is briefly described in the next sections.

### a) Hartree-Fock

The Born-Oppenheimer approximation is not enough, in a mathematical manner, to allow the energy calculation of a chemical system. Then, the Hartree-Fock theory (HFT) allowed the wave function calculation for any system, in an exactly manner, making the computer resources the major limitation. HFT describes the orbitals where the electrons move around, and the total wave function is given as a product of all those orbitals. [151] Nevertheless, these theories also introduce an approximation because they neglect the electronic repulsion between two electrons, considering the electron movement as a consequence of the electrostatic field created by the nuclei and the average field of the n-1 electrons (other electrons) in the system. Then, the electron movement could also be described by a single-particle function, being implicitly dependent on the motion of the other electrons. [152] Each single-particle function is assigned to each electron in the system and the wave function could be achieved by the product of all those functions (**Equation 30**).

$$\Psi_{Hartree\ Product}(r_1, \dots, r_n) = \phi_1(r_1) \times \dots \times \phi_n(r_n)$$

**Equation 30 – The Hartree product between the spatial orbitals of each electron in the system.**

Due to the Pauli exclusion principle, electrons with the same spin could not occupy the same orbital, so a spin orbital is calculated by the product between the spatial orbital and the spin function ($\alpha$ or $\beta$) (**Equation 31**).

$$\chi(x) = \phi(r).\alpha$$

**Equation 31 – Spin-orbital calculation by the product between spatial orbital ($\phi(r)$) and spin function ($\alpha$).**

The Hartree product is calculated using the spin orbitals instead of the solely spatial orbitals (**Equation 32**). [152]

$$\Psi_{Hartree\ Product}(x_1, \dots, x_n) = \chi_1(x_1) \times \dots \times \chi_n(x_n)$$

**Equation 32 – The Hartree product between the spin orbitals of each electron in the system.**

This calculation is not general, and it cannot be applied to all systems. Therefore, the spin orbitals determinant, called Slater determinant, should be used in order to normalize the calculation and define the expression for all systems (**Equation 33**).

$$\Psi_{electronic} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \cdots & \chi_N(x_1) \\ \vdots & \ddots & \vdots \\ \chi_1(x_N) & \cdots & \chi_N(x_N) \end{vmatrix}$$

**Equation 33 – Slater determinant**

Due the fact of no many-body system have a totally correct solution, the variational principle provides a mechanism that allows HF calculations give an approximated solution but always greater than the real value. This is important because the better the wave function, the lower is the energy, and consequently the energy minimum corresponds to the best wave function. [149]

Again, we can apply the variational principle to this type of wave function (Slater determinant) and obtain the following equation (**Equation 34**).

$$E_{electronic} = \int \chi_i \sum_{i=1}^{N} \widehat{h_\iota} \chi_i \, d\tau_i + \sum_{i=1}^{N} \sum_{j>i}^{N} (J_{ij} + K_{ij})$$

$$\widehat{h_\iota} = \hat{T}_{elec.} + \hat{V}_{elec.-nucleus}$$

**Equation 34 - Energy calculation using the HF theory.**

The last term in **Equation 34** corresponds to the bi-electronic interactions, and the $J_{ij}$ is the Coulomb operator, which represents the potential energy of an electron within the full electronic field distributed among the occupied spin orbitals. $K_{ij}$ is the exchange operator and represents the energy variation when two electrons exchange between two different orbitals. [152] The **Equation 34** is the sum of each spin orbital energy $\varepsilon_i$ obtained through the Fock operator ($\hat{F}$) (**Equation 35**).

$$\hat{F}\chi_i = \varepsilon_i \chi_i$$

**Equation 35 – HF equation for a spin orbital $\chi_i$.**

### b) Hartree-Fock Self-Consistent Field (SCF)

The Hartree-Fock equations seem to have a problem because we need to know all other spin orbitals to calculate the energy of a certain electron in the system. This issue is recurrent in this type of calculation, and it could be solved using the iterative

calculation to achieve a solution when the solutions are self-consistent. This method is called the self-consistent field (SCF). Recurring to this technique, a first trial set of spin orbitals is constructed based on the initial molecular structure. Those spin orbitals are used to build the Fock operator which is solved using the HF equations. If the convergence criterion is satisfied, the calculation ends and the system energy is calculated otherwise the new set of spin orbitals is used to construct a new Fock operator and the calculation restarts. [152]

### c) Linear Combination of Atomic Orbitals (LCAO)

HF calculations were successfully used when applied to atomic systems. However, the molecular application is computational inappropriate. Roothaan and Hall suggested an approximation that answered this problem, the linear combination of basis sets. This method described the molecular orbitals as functions of a solely electron centering in the atomic nucleus. The calculation is performed recurring to the linear combination of atomic orbitals theory (LCAO) which s mathematically decoded by **Equation 36**. [152]

$$\psi_{MO} = \sum_{r} c_r \chi_r$$

**Equation 36 – Molecular Orbital wave function based on a set of basis set using the spin orbitals ($\chi_r$) and associated coefficients ($c_r$).**

Theoretically, we can use an infinite number of basis functions to increase the accuracy, but we have to achieve a commitment between accuracy and computational cost.

### d) Semi-empirical

The high accuracy inherent to the HF calculations makes their application prohibited due to the high associated computational cost. The semi-empirical simplifies this issue through the use of a certain number of parameters that make the bi-electronic interactions calculation negligible, leading to fast calculations and allow their application in large systems. [152] The parameters could be obtained by high accurate computational methods or experimental data. However, it is a difficult to obtain these values because of the time-consumption of the great accurate methods and because of the low accurate results obtained by experimental methodologies. [149]

Semi-empirical calculations undertake the only explicit consideration of the valence electrons, correcting the atomic charge according to the cerne electrons. The major

difference between all the semi-empirical methodologies is derived to the neglect of differential overlap. The most primitive approach is called complete neglect of differential overlap (CNDO) which consider only the valence orbitals using the Slater-type of orbitals (STO), one function per orbital, all the two-electron integrals are parameterized and other simplifications to the HF theory are considered, reducing the two-electron integrals dependency from $N^4$ to $N^2$ and calculating the remaining through algebraic formulae instead of explicit integration. However, this formalism carries a non-negligible chemical cost leading to an incapable accuracy to predict good molecular structures. The intermediate neglect of differential overlap (INDO) was the successor of the CNDO formalism, and it aimed a better description of the electron-electron interactions based on the spectroscopic transitions. Then, a new formalism was considered, the neglect of diatomic differential overlap (NDDO), where the neglect of differential overlap is only applied when the basis function of a certain electron belongs to different atoms. Based on this formalism, several methods were developed such as MNDO (Modified Neglect of Diatomic Overlap), AM1 (Austin Model 1), PM3 (Parameterized Model 3) and PM6 (Parameterized Model 6). [149]

#### *i. Parameterized Model 6*

The PM6 method is a recent NDDO modification that was parameterized for 70 elements. This method presents an average unsigned error of only 4.4 $kcal.mol^{-1}$ when applied to 1373 compounds containing H, C, N, O, F, P, S, Cl, and Br. It was a significant improvement over the AM1 and PM3 methods. [153] Besides the interesting results that PM6 revealed, the molecular geometries predicted by this theory are less accurate than HF or DFT theories, for example. Nevertheless, the parameterization process itself carries an inherent error by the human choice of which parameters and data should be used to define the parameters.

### *2.4.1.2. Density-Functional Theory*

As it was mentioned before, the Schrödinger equation gives an exact solution for any system. However, its resolution is impossible for several relevant systems like enzymes. The HF considerations to simplify the calculations are also too heavy to be used in large poly-electronic systems because the requirement for the spin and three-dimensional coordinates for each electron. Furthermore, the wave function, *per si*, is hard to understand once it does not culminate in a physical property. [149] The semi-empirical attempt to solve the problem also does not offer a really good alternative to describe the molecular geometries accurately.

The Density-Functional Theory (DFT) described another way to predict the system's energy, for example, using easily concepts and mathematically-less complex approaches. In a general overview, a Hamiltonian depends only on the position and atomic numbers of the nuclei and the total number of electrons. In this manner to look at a chemical system, the electron density ($\rho$) is an observable physical property whose integration over all occupied space gives the total number of electrons (**Equation 37**). [149]

$$N = \int \rho(r) dr$$

**Equation 37 – Number of electrons ($N$) given by the integral of the electron density ($\rho(r)$) over all space.**

The electron density concept was introduced by Thomas and Fermi in the 1920s as a manner to simply the mathematical consideration of a function per each electron in the system. [154] [155]

Because the nuclei are in fact point charges, they could be described as local maxima in the electron density, appearing as point charges among the system's space. Now, the nuclear atomic number is the unique factor that was not included yet to build the Hamiltonian. This system property could then be easily introduced in the electron density. A certain nucleus A located at an electron density maximum ($r_A$) could also be introduced according to the **Equation 38**. [149]

$$\left.\frac{\delta \bar{\rho}(r_A)}{\delta r_A}\right|_{r_A=0} = -2Z_A\, \rho(r_A)$$

**Equation 38 – Electron density calculation for an atomic nucleus A with atomic number $Z_A$ and a spherically averaged density $\bar{\rho}(r_A)$.**

These concepts only make a Hamiltonian build possible using a certain electron density function, and it also presents an alternative manner to solve the Schrödinger equation. [149] The energy could be therefore calculated, using DFT, as a functional of the electron density, denoted $E[\rho]$. It means that the energy is given by a function where its arguments are another function, in this case the electron density. [152]

Furthermore, the DFT calculations are much faster than *ab initio* methods, when applied to systems with 100 or more atoms. Additionally, DFT calculations on d-block elements containing systems present results which agree more closely with experimental data comparatively with *ab initio* methodologies. DFT calculation is also faster than *ab initio* because of its dependency on only three coordinates (x, y and z coordinates of

$\rho(r)$) contrasting with *ab initio*, which depends on four variables per each electron (x, y and coordinates plus the spin value). [152]

The DFT energy calculation can be given by the **Equation 39** and the energy associated with internuclear repulsion is added after, according to Born-Oppenheimer approximation.

$$E_{electronic} = T[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)]$$

$T[\rho(r)]$ is the electrons' kinetic energy. $V_{ne}[\rho(r)]$ and $V_{ee}[\rho(r)]$ are the potential energy associated with nucleus-electron attraction and electron-electron repulsion, respectively.

**Equation 39 – Electronic energy calculation according to Thomas and Fermi.**

The potential energy terms are covered by a classical approach following the **Equation 40** and **Equation 41**.

$$V_{ne}[\rho(r)] = \sum_{a=1}^{M} \int \frac{Z_A}{|r - r_a|} \rho(r)\, dr$$

**Equation 40 – Potential energy calculation associated with the nucleus-electron attraction.**

$$V_{ee}[\rho(r)] = \frac{1}{2} \int \int \frac{\rho(r_i)\rho(r_j)}{|r_i - r_j|} dr_i\, dr_j$$

**Equation 41 – Potential energy calculation associated with the electron-electron repulsion.**

The kinetic energy calculation of a continuous charge distribution is conceptually difficult to understand. "Jellium" was introduced to answer this issue, which is a system where an infinite number of electrons are moving in an infinite volume of space that is characterized by a uniformly distributed positive charge. [149] In 1927, Thomas and Fermi projected an equation that allows the kinetic energy calculation for the "jellium" system (**Equation 42**) using fermion statistical mechanics.

$$T[\rho(r)] = \frac{3}{10} (3\pi^2)^{\frac{2}{3}} \int \rho^{\frac{5}{3}}(r) dr$$

**Equation 42 – Kinetic energy ($T[\rho(r)]$) calculation for the "jellium" system.**

This was an important historic step in the DFT development, but these considerations do not give better results than previous methods. Later on, Dirac proposed the addition of an exchange term (**Equation 43**) to **Equation 39**. [149] This exchange term allows the consideration of the Pauli's principle.

$$E_{exchange}[\rho(r)] = -\frac{9\alpha}{8}\left(\frac{3}{\pi}\right)^{\frac{1}{3}}\int \rho^{\frac{4}{3}}(r)\,dr$$

**Equation 43 – Exchange energy. $\alpha$ could be 1 if the Slater's derivation is considered. Some empirical values were attributed to $\alpha$.**

Still, this correction was not enough to make DFT a better alternative to previous methods.

### a) Hohenberg-Kohn Theorems

The DFT methodology was established and legitimated by the Hohenberg-Kohn theorems, in 1964, from which the first theorem asserts that in a non-degenerated system, the ground-state energy could be solely defined by the electron density. [156] Considering it, they proved that knowing the nuclei's positions (the external potential) and the number of electrons, the Hamiltonian is known, and consequently the wave function associated with the ground-state is also known.

The second theorem also called the Hohenberg-Kohn variational theorem, proved that the electron density followed the variation principle. Assuming a certain candidate density, the energy expecting value can be predicted according to the **Equation 44**. [149]

$$\langle \Psi_{candidate} | \hat{H}_{candidate} | \Psi_{candidate} \rangle = E_{candidate} \geq E_0$$

**Equation 44 – Energy calculated through a candidate density function is always greater than or equal to the ground-state energy.**

As described before, we could choose several density functions, and those that provide lower energies are closer to the real ground-state energy.

Besides the excellent assumptions of these two theorems, they do not give a real idea about the criteria that should be used to achieve the best density functions. [149]

### b) Kohn-Sham Self-Consistent Field (SCF)

In 1965, Kohn and Sham presented a way to achieve the system's energy calculation recurring to the density function. They proposed a hypothetical system where the electrons do not interact with each other whose Hamiltonian could be given by the sum of mono-electronic operators. According to this assumption, the electron density of the system is identical to the ground-state density and the energy could be given by the **Equation 45**. [157]

$$E[\rho(r)] = T_{ni}[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)] + \Delta T[\rho(r)] + \Delta V_{ee}[\rho(r)]$$

$T_{ni}[\rho(r)]$ is the kinetic energy of the non-interacting electrons. $\Delta T[\rho(r)]$ is the correction to the kinetic energy deriving from the interacting nature of the electrons. $\Delta V_{ee}[\rho(r)]$ is all the non-classical corrections to the electron-electron repulsion energy.

**Equation 45 – Energy calculation according to Kohn-Sham SCF theory.**

Considering this non-interacting system, the total kinetic energy is given by the sum of all individual electronic kinetic energies. The calculation is difficult since it requires the calculation of $\Delta T[\rho(r)]$ and $\Delta V_{ee}[\rho(r)]$ terms whose sum is commonly referred as exchange-correlation energy ($E_{XC}$). An orbital expression for the density may be written as the following equation:

$$E[\rho(r)] = \sum_{i=1}^{N}\left(\langle\chi_i| -\frac{1}{2}\nabla^2{}_i|\chi_i\rangle - \langle\chi_i|\sum_{a=1}^{M}\frac{Z_k}{|r_i - r_k|}|\chi_i\rangle\right) + \sum_{i=1}^{N}\left(\langle\chi_i|\frac{1}{2}\int\frac{\rho(r_j)}{|r_i - r_j|}dr_j\,|\chi_i\rangle\right) + E_{XC}[\rho(r)]$$

**Equation 46 – Energy calculation considering orbitals.**

This equation could be used together with the variational principle to predict the best density function to describe the system, in other words, the closest density to the ground-state one. As HF equations, the Kohn-Sham equations could be presented in a similar way of HF equations (**Equation 47**). [149]

$$h_i^{KS}\chi_i = \varepsilon_i\,\chi_i$$

**Equation 47 – Kohn-Sham equation for the orbitals $\chi_i$.**

$h_i^{KS}$ is the mono-electronic operator and is defined by:

$$h_i^{KS} = -\frac{1}{2}\nabla^2{}_i - \sum_{a=1}^{M}\frac{Z_k}{|r_i - r_k|} + \int\frac{\rho(r_j)}{|r_i - r_j|}dr_j + V_{XC}$$

$$V_{XC} = \frac{\delta E_{XC}}{\delta\rho(r)}$$

**Equation 48 – Kohn-Sham (KS) mono-electronic operator. $V_{xc}$ is a functional derivative.**

### c) Exchange-correlation functionals

As described above, DFT is an exact theory which needs a technique to get a good exchange-correlation energy approximation ($E_{XC}[\rho(r)]$). This functional may be

split into two ones, the exchange functional ($E_X[\rho(r)]$) and the correlation functional ($E_C[\rho(r)]$). The exchange functional is related to interactions between electrons with the same spin whereas the correlation functions are associated with the interaction between electrons of the opposite spin. Both functionals could be calculated using one of the two existing functionals, the local density approximation (LDA) and the generalized gradient approximation (GGA). [149]

The LDA is the simplest functional, and the electronic density is treated as a uniform distribution of electrons. This functional presented good results for systems containing metals, but it is inefficient for systems containing many different molecules because LDA requires a well-defined electronic density. LDA considers that each electron feels the same field regardless its position. [152]

The GGA circumvents this limitation through the density gradient inclusion, which takes into account the density variations throughout the system. Almost all GGA functionals are based on LSDA (Local spin-density approximation), a LDA-related functional which includes the spin. The exchange-correlation energy can be calculated, through GGA ($E_{XC}[\rho(r)]$), adding a density gradient-dependent term to the LSDA (**Equation 49**).

$$E_{XC}[\rho(r)] = -2^{\frac{1}{3}}C_x \int \left(\rho_\alpha^{\frac{4}{3}}(r) + \rho_\beta^{\frac{4}{3}}(r)\right) dr + \left(\frac{|\nabla\rho(r)|}{\rho(r)^{\frac{4}{3}}}\right)$$

**Equation 49 – Exchange-correlation energy calculation using GGA**

The first widely used exchange functional, named “B”, was developed by Axel Becke and his mathematical approach includes only one empirical parameter optimized considering the noble gases. A few more exchange functionals were developed based on B, such as CAM (Coulomb-attenuating method) [158], FT97 (developed by Filatov and Thiel in 1997) [159] and PW (Perdew-Wang) [160]. Other functionals were built using no empirical parameter like B88 [161], LG (Lacks and Gordon) [162] and PBE (Perdew–Burke–Ernzerhof) [163].

The correlation functional could be corrected using B88 [161], P86 (Perdew 86) [160] and PW91 [164] which do not include any parameter or LYP (Lee, Yang and Parr) [165] which includes four parameters but it is able to totally cancel the self-interaction error in one-electron systems.

These methods present the advantage of allowing the combination between any exchange functional with any correlation functional. BLYP is a popular example of that

combination once combines the Becke exchange functional with the Lee-Yang-Parr correlation functional. [149]

Later on, hybrid functionals were developed recurring to Hartree-Fock exchange functionals that are considered together with exchange and correlation DFT functionals, according to the following equation. [149]

$$E_{XC}^{Hybrid} = c_{HF}\, E_X^{HF} + c_{DFT}\, E_{XC}^{DFT}$$

**Equation 50 – Hybrid exchange-correlation energy calculation. $c_{HF}$ and $c_{DFT}$ are constants obtained by mathematical calculations fitting experimental data.**

B3LYP is the most used hybrid functional and combines the three parameters Becke functional (B3) and the LYP functional with some HF terms. The exchange-correlation energy calculated through the B3LYP functional follows the **Equation 51**. [149]

$$E_{XC}^{B3LYP} = (1-\alpha)E_X^{LSDA} + c_0\, E_X^{HF} + c_x\, \Delta E_X^{B88} + (1-c_c)E_C^{LSDA} + c_c E_C^{LYP}$$

**Equation 51 – B3LYP exchange-correlation energy calculation. $c_0$, $c_x$ and $c_c$ are empirical parameters determined by Becke.**

This functional provides the best results comparing with other exchange-correlation functionals, and the energy calculation error is about 3 kcal/mol when compared to true *ab initio* methods. [149]

Recently, in 2005, Minnesota functionals, developed by Truhlar, introduced a higher weight for HF exchange energy. The first functional, M05 [166], considered 28% of HF exchange energy while the successor M05-2X [167] includes 56%. In 2007, the M06 family of functionals [168], M06 including 27% of HF correction and M06-2X with 54% were published, and they appear as an improvement of the previous M05 functionals. The mean error of energy calculation decreased comparing with M05 and allow a more accurate energy determination. The exchange-correlation energy calculation follows:

$$E_{XC}^{M06} = \frac{X}{100} E_X^{HF} + \left(1 - \frac{X}{100}\right) E_X^{DFT} + E_C^{DFT}$$

$X$ is equal to 27 for M06 functional and 54 for the M06-2X.

**Equation 52 – Exchange-correlation energy calculation used by the M06 family of functionals.**

Unfortunately, M06 methods require highly computational resources that make its use inadequate when biological systems are studied. Consequently, in this work, the

M06 methods are used to calculate the energies in a more accurate manner, but the geometry optimization should be done recurring to the B3LYP functional.

#### d) Dispersion corrections

The major disadvantage of DFT methods is the poor description of the London dispersion interactions. These interactions are very important for large systems (e.g. enzymes) in order to reach an excellent chemical accuracy. The dispersion correction functional (DCF) is an add-on to DFT calculation once it allows an energy calculation refinement achieved through a better description of the dispersion interactions. The most recent parameterization of DCF is DFT-D3, and it was parameterized by the Grimme's group. [169] The DFT-D3 functional was improved by Becke and Johnson (BJ) through the introduction of an extra damping functional. The damping functional does not require any pair-specific cut-off radii, and it also cancels all artificial repulsive interatomic forces at short distances. Besides the better description, no significant differences were observed in the energies calculated using the BJ functional or the original DFT-D3 functional. [170] Furthermore, there are not parameters to use BJ functional as an add-on functional of M06 or M06-2X. Consequently, in this work, only the DFT-D3 functional was used to correct the London dispersion interactions.

### *2.4.1.3. Basis Sets*

Basis sets are sets of mathematical functions used to describe atomic orbitals in quantum mechanical calculations as DFT and *ab initio*, in contrast with parameter-dependent methods such as semi-empirical and molecular mechanics. The results are much better as more accurate are the orbital description by the mathematical functions. In the limit, we could use an infinite number of basis sets in order to describe orbitals more accurately as possible. However, it has a computational cost, and a compromise should be taken into account between accuracy and computational cost. Consequently, two main criteria must be considered when a basis set limitation is assigned: the basis set must be able to describe the real wave function and generate useful chemical results; furthermore, those sets should offer a simple way to solve the exchange-correlation functionals without an excessive computational cost. [149]

#### a) Slater-type orbitals (STO)

The Slater-type orbitals (STO) functions were the first to be applied, and they present an exponential dependence of the nucleus-electron distance, describing the

hydrogen atomic orbitals accurately. This exponential approach allows a fast convergence process. However, their application is limited to systems with a few number of atoms due to the difficulty to calculate their four center integrals. STO orbitals do not include the orbital radial nodes, but they could be introduced through the linear combination of STOs. [149, 152]

### b) Gaussian-type orbitals (GTO)

Boys introduced, in 1950, an improvement to STO replacing the $e^{-r}$ dependence of the STO functions by a $e^{-r^2}$. This was the turning point and the establishment of the Gaussian-type orbitals (GTO) defined as:

$$\chi^{GTO}(x,y,z;\alpha,i,j,k) = \left(\frac{2\alpha}{\pi}\right)^{\frac{3}{4}} \left[\frac{(8\alpha)^{(i+j+k)} i!\, j!\, k!}{(2i)!\,(2j)!\,(2k)!}\right]^{\frac{1}{2}} x^i y^j z^k e^{-\alpha(x^2+y^2+z^2)}$$

$\alpha$ determines the GTO width. $i, j, k$ are the non-negative integers that define the orbital nature. $x, y, z$ are the atom position coordinates.

**Equation 53 – GTO function in atom-centered Cartesian coordinates.**

If all defining-orbital integers are zero, the GTO is an s-type of orbital. In a case where one of the integers gets the unitary value, a p-type GTO is defined, and we could have three combinations corresponding to $p_x$, $p_y$ and $p_z$. The d-type of orbitals are six, and they are defined by the $i, j, k$ sum equal to two, leading to a six Cartesian prefactors of $x^2$, $y^2$, $z^2$, $xy$, $xz$ and $yz$. The Schrödinger equation only predicts the first five, being the two of the last ones obtained by linear combination of orbitals. In fact, the remaining one is an s-type of orbital due to its spherical symmetry. [149]

Besides the efficient mathematical calculation allowed by GTO functions, they do not reproduce the orbital amplitude accurately along the nucleus-electron distance because their $r^{-2}$ dependence. Consequently, the contracted GTO functions (CGTO) were developed through the linear combination of STO and GTO orbitals aiming a correct orbital description with a low computational cost achieved with GTO functions. The CGTO functions improve the description of the electrons near the atomic nucleus, but due to their insignificant role in the chemical process, this approximation does not limit the results quality. [149, 152]

Later on, Pople introduced the split-valence basis concept where the core orbitals are treated under a certain scheme of Gaussian primitives whereas the valence orbitals are treated using two or three different schemes of Gaussian contraction. A nomenclature method was also suggested to describe this method: *n-abc*G. *n* represents

the number of primitives used to contract GTO of the core orbitals, and the *abc* represents the division and the treatment assigned to the valence orbitals. For example, a common basis set 6-31G is characterized by six primitive GTOs for the core orbitals and the valence orbitals are divided into two from which the most internal is described through three primitive GTOs and the external ones use only one. [149, 152]

CGTOs can describe accurately isolated atoms, but the description is poor when molecular systems are considered because they are not polarized and their shape is not modulated by the surrounding atoms. Pople also introduces the polarization and diffuse functions, which can be combined with previous GTOs, to represent the interactional phenomena. [149, 152]

The polarization functions are characterized by a higher angular momentum of the valence orbitals. Considering the hydrogen atom, a p-type orbital can be used to polarize the s-type orbital of the hydrogen atom, distorting their electron cloud defining the chemical bond. We could also have d-orbitals polarizing p-orbitals which are an extremely relevant consideration for the second-period elements. Pople also defined a nomenclature to named this polarization functions. If we have a 6-31G set that uses d-orbital polarization, they could be written as 6-31G(d) or 6-31G*. Additionally, the s-orbitals polarization by p-orbitals could be included and the set is named 6-31G(d,p) or 6-31G**. In this work the 6-31G(d) basis set was used to perform the geometry optimization due to their good balance between computational cost and accuracy. [149, 152]

The diffusion functions are essential for the correct description of highly negative molecules where the electron cloud is considerable more displaced relatively to the nucleus. This functions includes a different exponential and are denoted by “+” or “++” if they assigned the diffusion functions to heavy atoms or heavy and hydrogen atoms, respectively. Considering a system where the hydrogens and heavy atoms are considered using diffusion functions, the basis set is written 6-31G++(d). [149, 152]

### 2.4.2. Nuclear Hamiltonian

All the previous calculation methods are able to determine solely the electronic energy of a system plus the energy that derives from the interaction between the nucleus and the electrons (electronic Hamiltonian). However, in order to obtain the full energy, it is necessary to calculate the kinetic and potential energies from the nucleus. This energy is calculated using the nuclear Hamiltonian and uses certain approximations to achieve its calculation.

The calculation of the nuclear Hamiltonian is extremely important when the catalytic mechanisms are studied because it provides the activation and reaction $\Delta G$ (free Gibbs energy variation). Besides the slight effect of the nuclear movement into the global system energy, its effect could be a cause of the disagreement between experimental determinations and theoretical calculations. Even at 0 K, the nuclear motion is not null, and the internal energy of the system is given by the sum of the electronic energy and the zero-point energy (ZPE). Further, the thermal corrections are computed and added to the electronic energy in order to obtain an energy that fits a certain temperature and allows the comparison with experimental data. The internal energy at a certain temperature could be calculated by the following equation:

$$E_{T=273.15\,K} = E_{electronic} + [E_{translational} + E_{rotational} + E_{vibrational}]_{T=273.25\,K}$$

**Equation 54 – Internal energy calculation at a certain temperature.**

The last three terms of the **Equation 54** compose the nuclear Hamiltonian and they are calculated classically being the translational energy ($E_{translational}$) calculated under the particle in a box model, the rotational energy ($E_{rotational}$), calculated based on rigid rotor model and the vibrational energy ($E_{vibrational}$) that is computed for each vibrational mode of a molecule.

Additionally, the normal mode analysis of the vibrational energies provides the values of all the frequencies in the system. These frequencies are calculated through the second derivative of the energy. Commonly, when we study a catalytic mechanism, we are interested in the minima (reactant and product) and the TS. The normal mode analysis allows the TS characterization, once a TS has one and only one imaginary frequency. The imaginary frequency must involve the atoms that are related with the reaction coordinate. Otherwise, both minima have only positive frequencies.

# 3. Hybrid Methods

The previous sections describe different theoretical methods characterized by different accuracy and consequently different computational cost. Obviously, the QM-based methods, as DFT or *ab initio*, present a high exactness in the chemical geometry and energy description. However, their application is limited above a hundred atoms because their high computational requirements. When we want to study biological systems, such as enzymes, we have to consider a huge number of atoms, habitually thousands of them. QM methods cannot be considered, and only MM methods could be used to study these systems. Unfortunately, MM does not consider electrons and its classical approach is not able to describe chemical reactions involving the bond breakage and formation.

In last few years, the hybrid methods have been considered and developed to assign different theoretical levels to different parts of a certain system. Those hybrid methods are commonly called QM/MM because the most considered a small part of the system using a QM-based method and the remaining through MM. That approximation is acceptable because usually the electron-involving phenomena are constricted to the active site. A careful choice of which atoms should be studied using a QM method leads to very good results and an accomplishable computational calculation. [149, 151, 152]

The present work uses precisely the QM/MM methodology to address the mHDC catalytic mechanism, using DFT to describe the active site relevant atoms, being the remaining protein treated under MM approach.

## 3.1. Type of methods

The first step that should be taken, when a QM/MM methodology is used to study a certain system, is the choice of which atoms will be studied using which method. Commonly, biological systems are studied considering the relevant active site amino acid residues, substrates and/or co-factor in the high-level layer (HL) whose calculation is done using a QM method. The current computing power does not allow more than few hundreds of atoms in the HL. The remaining entire protein and solvent correspond to the low-level layer (LL), and it should be treated by MM methods. The big challenge of this hybrid approach is related to the boundary zone and how could the connection between layers be made. If the bond-interactions appears to be difficult to consider, the non-bonding interaction, as van der Walls and electrostatic ones, is much more complicated

due to the polarization that atoms from one layer could induce in the other. Techniques to address this problem will be explained ahead.

The energy system calculation can be done using mainly two types of approaches, the additive and subtractive methods, in order to combine the energy calculated to each layer. In this work, the subtractive method was used because the additive method has not accurate equation and parameters to efficiently calculate the coupling term of its energy equation. [171]

Considering a two region split system (HL and LL), three calculations should be done to calculate the total system energy ($E_{QM/MM}^{All}$). According to the **Equation 55**, first all the system is considered under the MM theoretical level ($E_{MM}^{All}$). Then, only the HL atoms associated energy is calculated using the MM method ($E_{MM}^{HL}$). The LL region energy is calculated subtracting this last energy to the first one. Finally, the energy of the HL atoms is computed again using the QM theoretical level ($E_{QM}^{HL}$) and the resulting energy summed.

$$E_{QM/MM}^{All} = E_{MM}^{All} - E_{MM}^{HL} + E_{QM}^{HL}$$

**Equation 55 – Energy calculation of a subtractive-type of hybrid method**

The system could be divided into more regions and the total energy calculated is using the same type of protocol.

### 3.1.1. ONIOM

ONIOM (our own *N*-layered integrated molecular orbital molecular mechanics) is one of the most used subtractive hybrid methods and allows a multiple system division. [172] According to the authors, a protein could be studied using MM to describe the external region; a semi-empirical method for the intermediate amino acid residues; and the active site region can be calculated using a DFT or *ab initio* method. ONIOM requires an explicit description of charges and spin multiplicities for the different regions. [149]

In this work, a two-layer system (**Fig. 18**) was used, and the HL calculation recurred to DFT (B3LYP) method, and LL used MM approach. The energy was calculated according to the following equation:

$$E_{DFT/MM}^{All} = E_{MM}^{All} - E_{MM}^{HL} + E_{DFT}^{HL}$$

**Equation 56 - Energy calculation of the two-layered ONIOM method.**

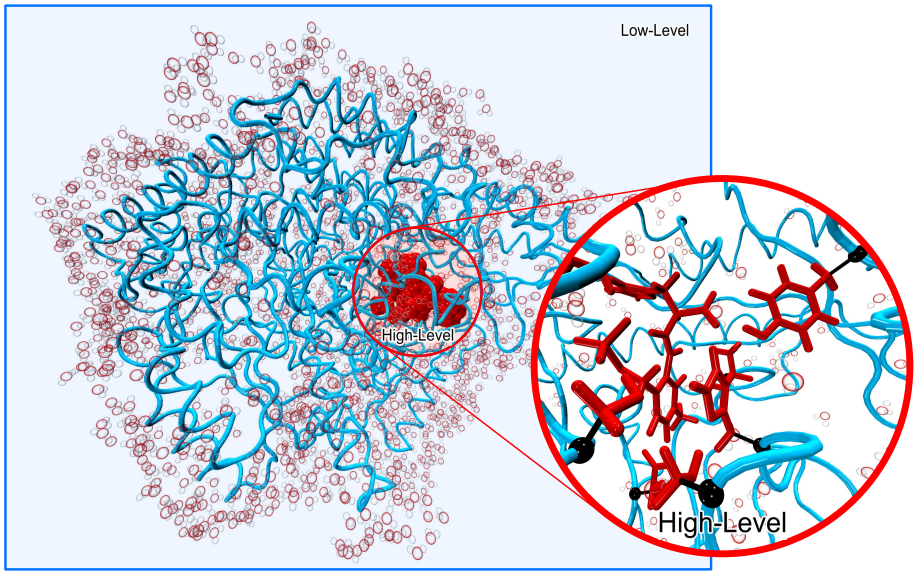



**Fig. 18 – Representation of the ONIOM two-layer method.**
Red spheres represent the high level atoms considered using DFT method. Water molecules (Glass bubble representation) and the remaining protein (blue ribbon representation) compose the low level treated by MM. Black spheres bonded to the high level atoms represent hydrogen atoms used under the link atoms approach.

## 3.2. Boundary methods

The hybrid method raises a problem related with the atom linkage between atoms belonging to different layers. The problem is quietly bigger when this interatomic interaction is covalent which is commonly founded in protein studies. Two approximations were developed in order to overcome this problem: the link atoms approach and the frozen orbital approach. [149]

### 3.2.1. Link atoms approach

Link atoms approach allows the structural and energetic correction due to the boundary cut between layers. This approach uses hydrogen atoms to bind directly to the HL atoms whose bonds were cut in the division of layers (**Fig. 18**). The hydrogen atoms can be replaced by other elements such as halogens or group of atoms. However, the added atoms are not considered when the MM calculation is computed because those atoms are not part of the system. The energy of the bonds which were cut are evaluated by MM method considering the respectively HL atom as a LL one. Some studies neglect a few of intramolecular features such as angle and torsional bending. [149]

The Gaussian software adds the link atoms through a vector whose direction is dictated by the HL atom configuration. The link atom is placed at a certain typical distance which depends on the involved elements. [149]

Besides the link atoms approach simplicity, their application introduces some geometry and energy related errors due to the link atoms induced polarization on the HL. However, this is the most used approach mainly due to the difficulties inherent to frozen orbital approach implementation.

## 3.3. Communication between layers

In addition to the bonded atoms description, hybrid methods also present other problem related to the electrostatic and van der Waals interactions between atoms belonging to different layers. Consequently, polarization effects are neglected, and the boundary regions tend to be over polarized. Some methods were developed to transfer the polarization effect between layers. Nonetheless, their application could be more or less extensive, carrying a high computational cost. The mechanical and electrostatic embedding methods are the two most used. [149]

### 3.3.1. Mechanical embedding

The mechanical embedding method is the simplest one, allowing a LL polarization induced by the HL atoms. The HL region is calculated using the QM method without any concern about the polarization. Then, the LL region is calculated using MM method feeling the HL polarization, which is introduced through punctual charges. In fact, the charges of HL atoms are calculated in each step, but the system does not undergo a new optimization cycle concerning those new charges of HL. Habitually, Lennard-Jones parameters are not updated along the simulation. Some consequences can arise from that approximation, mainly the changes in hybridization states whose influence is not accounted for. Fortunately, Lennard-Jones potential is a quite short-range function introducing not very significant error.

### 3.3.2. Electrostatic embedding

The electrostatic embedding allows a HL polarization by the LL region, which is introduced through an additional mono-electronic term that is added to the QM Hamiltonian. This term is calculated from the atomic charge distribution along the LL region. Otherwise, the LL polarization due to the HL atoms is neglected in this approach, but the results tend to be considerably better than those obtained by mechanical embedding approach. In sum, electrostatic embedding scheme adds a new self-consistent cycle where the MM layer must converge according to the new orbital

coefficients determined after each cycle for the QM layer. The main disadvantage of this approximation is the computational effort increase.

This page was intentionally left blank.

# C. Results and Discussion

*Everything is theoretically impossible,*

*until it is done.*

**Robert Heinlein**

This page was intentionally left blank.

In this work, we studied the catalytic mechanism of mHDC using an atomistic perspective. The only crystallographic structure of the hHDC available on the Protein Data Bank (PDB) was used to study the catalytic mechanism.

# 1. Structure preparation

Since the only available structure has an inhibitor in the active site, instead of the natural substrate, we modified it accordingly. The following section will depict the structure of hHDC, and the procedures that were done to prepare the models.

The PDB file of hHDC includes three dimers that are co-crystallized with an inhibitor (HME), some structural modifications were done in order to get a representative model of hHDC with the natural substrate. The present section described the steps used to obtain that representative model.

## 1.1. PDB Analysis

First of all, only one of the dimeric structures of HDC was selected from the PDB file (PDB code: 4E1O). It is important to take into account the dimeric structure of hHDC since the active site of this enzyme (one in each monomer) is composed of residues of both monomers. The dimer that was chosen includes chains A and B even though Gly-1 is missing. The dimer containing chains C and D was not considered because the chain C has a Gly-Pro-Leu tag, which was artificially introduced to help the protein purification. The third dimer (including chain E and F) was also discarded because comparatively to dimer AB, it provides a worst orientation of the PLP-substrate inside the active site. (**Supplementary Fig. 1**)

In our models, each subunit of hHDC contains 477 amino acid residues, resulting in a total of 956 residues for the active dimer. Each subunit comprises fourteen α-helices, nine $3_{10}$-helices, five β-bridges and eleven β-sheets. [17] Active hHDC has also two active sites that are located in the interface region between the two subunits that comprise amino acid residues from both monomers, i.e., Asp-273A, His-194A, Ser-151A, Arg-447A, Ser-154A, Thr-248A and Ser-196A from chain A and Tyr-334B and Ser-354B from chain B (**Fig. 19**).

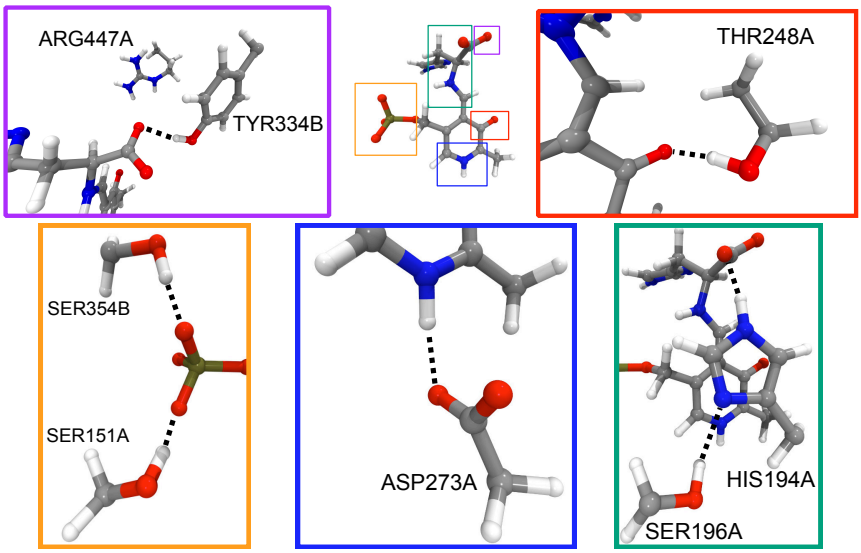



**Fig. 19 – Amino acid residues at the active site interacting with the external aldimine.**
CPK representation of the external aldimine (top center image) and CPK representation of the side chain of each amino acid residue labeled with name, number, and chain; and non-bond interactions (dash line).

## 1.2. Model Preparation

In the co-crystalized structure the HME inhibitor is already bonded to the PLP cofactor in the form of an external aldimine (**Fig. 20** Left). In order to model the "natural" external aldimine (**EA-S** model), the methyl group was removed from that structure (**Fig. 20** Right).

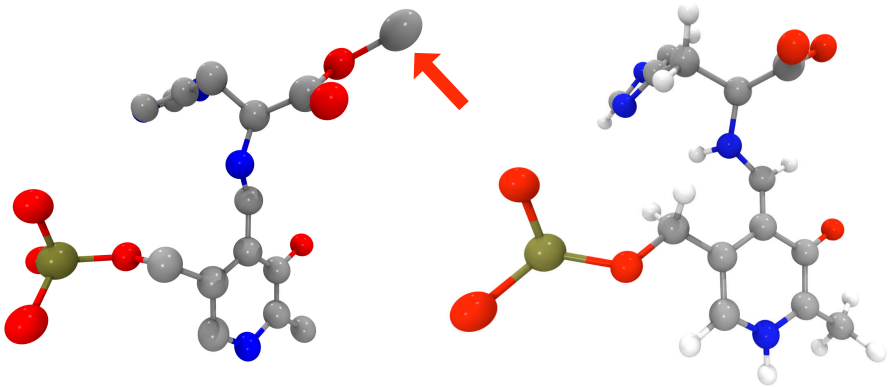


**Fig. 20 – External aldimine representation.**
CPK representation of (Left) the hHDC inhibitor (HME) bonded to the PLP cofactor. Structure from original PDB file (PDB code: 4E1O) without protonation; and (Right) the hHDC substrate (*L*-histidine) bonded to the PLP cofactor. Structure modified from the PDB file.

Experimental data showed that a quinonoid intermediate (QI) is formed during the conversion of *L*-histidine into histamine by hHDC. [26] Therefore, the **QI** model (**Fig.**

**21** Right) was also prepared by removing the carboxylate group from the external aldimine of the **EA-S** model (**Fig. 21** Left).

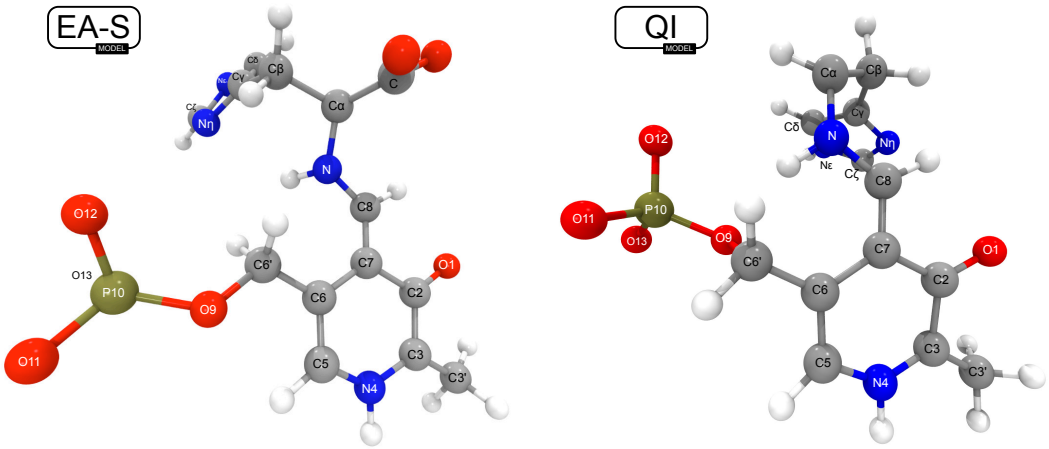



**Fig. 21 – CPK representation of (Left) the EA-S model and (Right) the QI model.**
An atom designation was assigned to each atom of the external aldimine and quinonoid intermediate.

As mentioned before, two cysteine residues (C180S and C418S) were mutated to help the crystallization process of the complex between the enzyme and the inhibitor. These residues were mutated to the proper amino acid residue, in the built model, as it would be found in wild-type enzyme. [17]

## 2. Molecular Dynamics Simulations

After the adequate preparation of the structures, MD simulations were run for the resulting models (**EA-S** and **QI** models). MD simulations allow to equilibrate the structure and get an ensemble of conformations from which we could get hints about the catalytic mechanism. Furthermore, they also generated relevant structural information such as: i) hydrogen bond interactions between the external aldimine or the quinonoid intermediate with the residues of the binding site; ii) the flexibility of some amino acid residues from the active site; and iii) conformational rearrangements of the external aldimine (PLP bonded to the substrate) in contrast with the quinonoid intermediate (PLP bonded to the decarboxylated substrate).

### 2.1. Parameterization

MD simulations use MM, which requires the parameterization of some molecules that are not fully parameterized on the available force fields. In this case, the external aldimine and the quinonoid intermediate do not have their parameters available in the AMBER force fields (ff99SB [135] and GAFF [173]). Therefore, the required atomic charges were obtained through an *ab initio* methodology, the Hartree-Fock (HF), with the 6-31G(d) basis set. Additionally, ANTECHAMBER program was used to assign the atomic typology of each atom from external aldimine and quinonoid intermediate (**Fig. 22**). The determined parameters are summarized in **Table 2**.

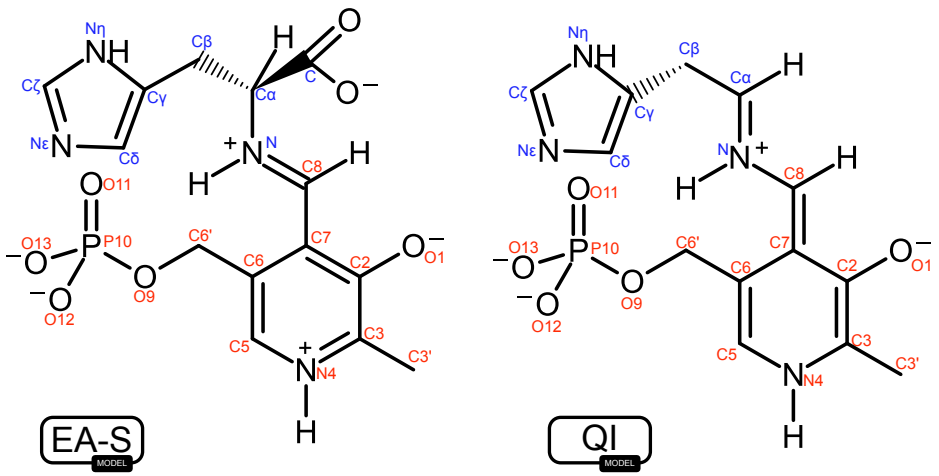



**Fig. 22 – Wedge-Dash representations of the (Left) external aldimine of EA-S model and (Right) the quinonoid intermediate of QI model, and respectively atom designation.**

**Table 2 - Atomic charges (atomic units) assigned to each model.**
Charges were obtained through *ab initio* (HF/6-31G(d)) calculation. Subscript notation identifies the atom to which a certain atom is bonded (e.g. $H_{N4}$ represents the hydrogen bonded to the nitrogen N4).

| Atom designation | Model EA-A | | Model Q-A | |
|---|---|---|---|---|
| | Charges | Atom Type | Charges | Atom Type |
| C | 0.767276 | c | -- | -- |
| C2 | 0.435869 | c | 0.238541 | c |
| C3 | 0.155984 | cc | 0.021528 | cc |
| C3’ | -0.385323 | c3 | -0.228211 | c3 |
| C5 | -0.158819 | cd | -0.076425 | cd |
| C6 | 0.024505 | cc | -0.176716 | cc |
| C6’ | 0.091057 | c3 | 0.243755 | c3 |
| C7 | -0.200725 | cc | 0.153613 | cc |
| C8 | 0.295049 | c2 | -0.217836 | cf |
| Cα | 0.227887 | c3 | 0.136627 | c2 |
| Cβ | -0.320786 | c3 | -0.186391 | c3 |

| | | | | |
|---|---|---|---|---|
| Cγ | 0.322029 | cc | 0.401509 | cc |
| Cδ | -0.300586 | cd | -0.243540 | cd |
| Cζ | 0.253494 | cd | 0.224967 | cd |
| $H_{C3'}$ | 0.110522 | hc | 0.034076 | hc |
| $H_{C3'}$ | 0.110522 | hc | 0.034076 | hc |
| $H_{C3'}$ | 0.110522 | hc | 0.034076 | hc |
| $H_{C5}$ | 0.228944 | h4 | 0.131444 | h4 |
| $H_{C6'}$ | 0.040947 | h1 | -0.010255 | h1 |
| $H_{C6'}$ | 0.040947 | h1 | -0.010255 | h1 |
| $H_{C8}$ | 0.143579 | h4 | 0.104414 | h4 |
| $H_{Cα}$ | 0.006068 | h1 | 0.075237 | h4 |
| $H_{Cβ}$ | 0.091222 | hc | 0.041361 | hc |
| $H_{Cβ}$ | 0.091222 | hc | 0.041361 | hc |
| $H_{Cδ}$ | 0.217632 | h4 | 0.247842 | h4 |
| $H_{Cζ}$ | 0.087502 | h5 | 0.077188 | h5 |
| $H_N$ | 0.353722 | hn | 0.375443 | ho |
| $H_{N4}$ | 0.373438 | hn | 0.308858 | hn |
| $H_{Nε}$ | 0.384845 | hn | 0.353390 | hn |
| N | -0.287710 | nh | -0.248950 | nf |
| N4 | -0.267891 | na | -0.512076 | na |
| Nε | -0.348962 | nc | -0.386777 | na |
| Nη | -0.661535 | nc | -0.655376 | nc |
| O1 | -0.691859 | o | -0.748082 | o |
| O11 | -0.907394 | o | -0.812165 | o |
| O12 | -0.907394 | o | -0.812165 | o |
| O13 | -0.907394 | o | -0.689602 | oh |
| O9 | -0.439082 | os | -0.471133 | os |
| $O_C$ | -0.779162 | o | -- | - - |
| $O_C$ | -0.779162 | o | -- | -- |
| P10 | 1.379001 | p5 | 1.206651 | p5 |

According to the parameterization, the release of carbon dioxide generates a negative charge into carbon Cα, which is delocalized along the PLP and accommodated by nitrogen N4. This is corroborated by decrease of the charge of nitrogen N4 between the **EA-S** and the **QI** models.

### 2.2. MD Simulations

The topology and parameter files required to run an MD simulation were prepared and built using the ANTECHAMBER software from the AMBER12 package. The parameters were retrieved from the ff99SB [135] and the General AMBER (GAFF) force fields [173]. The missing parameters were included taking into account the data that was obtained from the previous section. The total charge of each studied system was

neutralized with 18 sodium ions ($Na^+$) and solvated using a TIP3P [174] type of water molecules distancing at least 12 Å from protein (**Supplementary Fig. 2**).

After the structural minimizations, the system was submitted to an equilibration stage and subsequently to an MD simulation production, using the SANDER software from AMBER12 package [134]. During the 20 ps (integration step = 1 fs) of the equilibration stage, the temperature of the system was incrementally increased from 0.00 K to 310.15 K using the Langevin thermostat [175]. In this step, we considered periodic boundaries conditions and a canonical (NVT) ensemble. The non-bond interactions were considered up to the distance of 10.0 Å. Subsequently, a production stage of 60 ns was done considering an isothermal-isobaric (NPT) ensemble.

The MD simulation results were afterward visualized in the Visual Molecular Dynamics software (VMD) [176] and analyzed with Ptraj software [177] from the AMBER12 package [134].

### 2.2.1. System Equilibration

During the MD simulations, general physical properties of the system, such as kinetic and potential energies, temperature, volume, pressure and density, were evaluated for the two model systems. The obtained results indicate that the systems were stabilized, and no abnormal behavior is observed. Special attention was given to the pressure that showed a high fluctuation. Nevertheless, it is considered a normal situation because of the small dimensions of the system, and therefore little atomic movements leading to a high pressure oscillation. (**Supplementary Fig.3**, **Supplementary Table 1**, **Supplementary Fig.4** and **Supplementary Table 2**).

A further root mean square deviation (RMSd) determination was made to the MD simulations concerning about the conformation movements, and structural deformation of the protein. RMSd is a measure of atomic spatial displacement relatively to the initial position.

We verified that the MD simulation of the **EA-S** model did not cause a significant deformation of the protein structure, once the global RMSd slightly exceeded 2 Å (**Fig. 23**). Evaluating the RMSd values along the simulation, we can consider an equilibrated system between 40 and 60 ns. The equilibrated region presents a global RMSd of 1.78 ± 0.21 Å and a backbone RMSd of 1.31 ± 0.22 Å. Considering solely the active site residues (Asp-273A, His-194A, Ser-151A, Arg-447A, Tyr-334B, Ser-354B, Ser-154A, Thr-248A and Ser-196A) and the **EA-S**, the RMSd presents an even lower fluctuation,

1.06 ± 0.19 Å. Only the structures within the equilibrated system (40 to 60 ns) were studied in additional analysis and calculations.

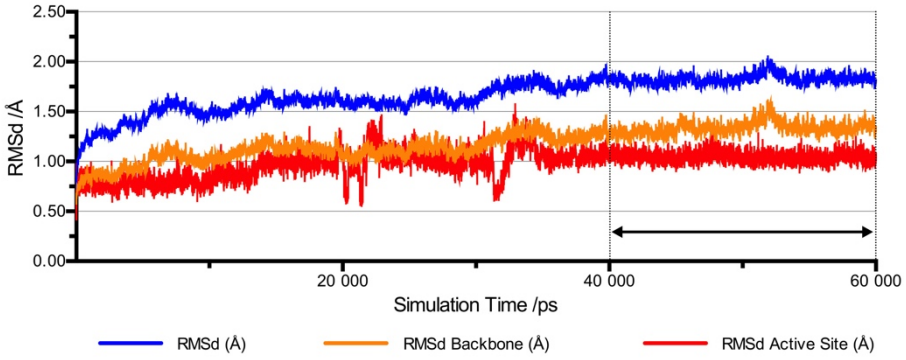



**Fig. 23 – RMSd values for the MD simulation of the EA-S model.**
The blue line corresponds to the RMSd value of all atoms in the system. The orange line corresponds to the RMSd values considering only the α-carbons, corresponding to the protein backbone. The red line corresponds to the RMSd values considering the active site amino acid residues and the external aldimine. The system was considered equilibrated during the last 20 ns (40 to 60 ns).

We also looked to the RMSd of each subunit in order to disclose if some abnormal fluctuation could be hidden by the overall mean. Nevertheless, no abnormal divergence was observed, and both subunits presented a similar behavior along the full simulation (**Fig. 24**).

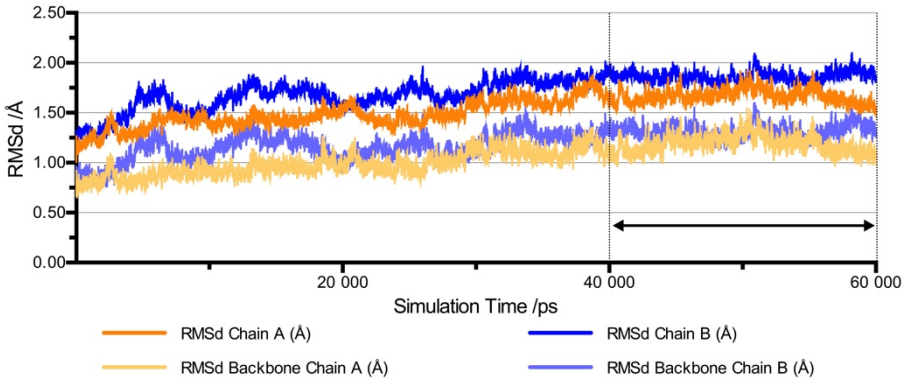



**Fig. 24 – RMSd values for the MD simulation of the EA-S model.**
The orange lines correspond to the RMSd values of the amino acid residues of the subunit A. The blue lines correspond to the RMSd values of the amino acid residues of the subunit B. The system was considered equilibrated during the last 20 ns (40 to 60 ns).

Similarly to the **EA-S** model, the MD simulation of the **QI** model did not show a significant structural displacement along the simulation comparing with the first structure. In the same way, the global RMSd values slightly exceed 2.5 Å (**Fig. 25**).

Again, we considered the last 20 ns, between 40 and 60 ns, where the structure is equilibrated. For that time lapse, the global RMSd value is 2.12 ± 0.30 Å in contrast

with RMSd value of backbone which is 1.73 ± 0.31 Å. Considering only the active site amino acid residues and the **QI**, the RMSd is even lower, 1.30 ± 0.25 Å.

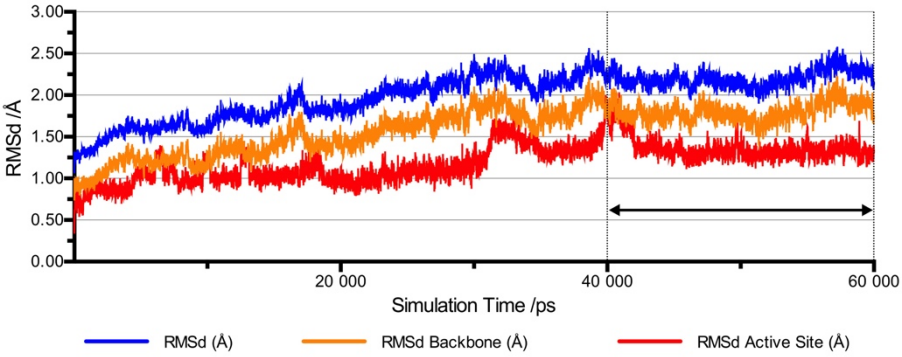



**Fig. 25 – RMSd values for the MD simulation of the QI model.**
The blue line corresponds to the RMSd value of all atoms in the system. The orange line corresponds to the RMSd values considering only the α-carbons, corresponding to the protein backbone. The red line corresponds to the RMSd values considering the active site amino acid residues and the quinonoid intermediate. The system was considered equilibrated during the last 20 ns (40 to 60 ns).

The RMSd evaluation for each subunit also revealed no significant differences between both subunits, showing a similar behavior (**Fig. 26**).

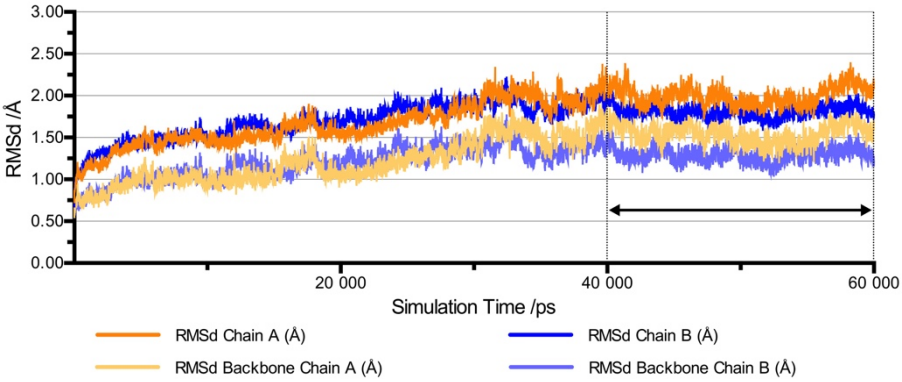



**Fig. 26 - RMSd values for the MD simulation of the QI model.**
The orange lines correspond to the RMSd values of the amino acid residues of the subunit A. The blue lines correspond to the RMSd values of the amino acid residues of the subunit B. The system was considered equilibrated during the last 20 ns (40 to 60 ns).

The small difference between the RMSd values of both simulations (**EA-S** and **QI** models) can be easily explained by the fact that the quinonoid intermediate of the **QI** model does not diverge so much from the external aldimine of the **EA-S** model. However, the **EA-S** model presents lower RMSd values once it fits more closely with the original PDB structure than the **QI** model which does not have the carboxylate group. The carboxylate group elimination induces, in the **QI** model, a small protein rearrangement targeting it to a new energy minimum. This new disposal of the active site requires more

movements from the surrounding atoms of the active site, resulting in the obtained slightly higher RMSd values for the **QI** model.

Further analysis will only concern the last 20 ns of each MD simulation, corresponding to an equilibrated system.

### 2.2.2. Structural Analysis

The following analyses will address a detailed evaluation of each model (**EA-S** and **QI**), concerning: i) their key differences relatively to the original PDB structure; ii) the important interactions between the active site residues and the **EA-S** and **QI** reaction intermediates; and iii) structural patterns.

#### *2.2.2.1. Root-Mean-Square deviation (RMSd)*

Comparing all the structures from the equilibrated time lapse of the **EA-S** model simulation (40 to 60 ns) with the original PDB structure, no significant conformation deformation was observed since the RMSd value is 1.86 ± 0.06 Å (**Fig. 27** Blue line). Considering only the backbone, the RMSd is even lower, 1.37 ± 0.08 Å (**Fig. 27** Orange line), and much lower if we only consider the active site residues and the external aldimine of the **EA-S** model, 0.83 ± 0.09 Å (**Fig. 27** Red line).

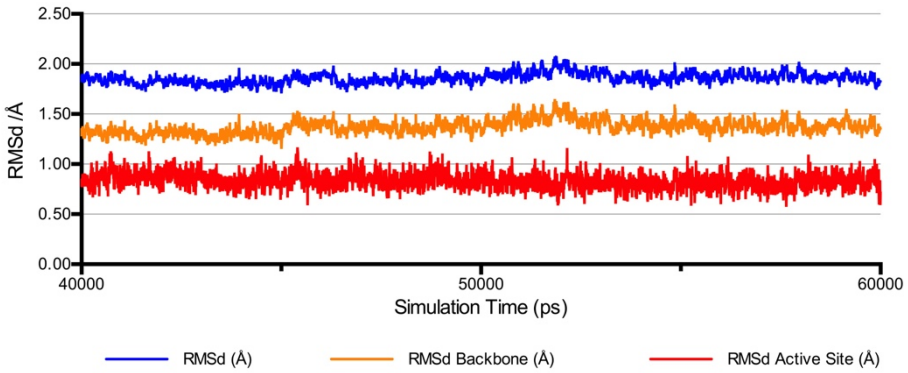



**Fig. 27 – RMSd values for the equilibrated MD simulation of the EA-S model using the original PDB structure as a reference.**
The blue line corresponds to the RMSd value of all atoms in the system. The orange line corresponds to the RMSd values considering only the α-carbons, corresponding to the protein backbone. The red line corresponds to the RMSd values considering the active site amino acid residues and the external aldimine.

Overlapping the two structures, we can verify that no observable changes happen during the MD simulation (**Fig. 28** Left). The residues from the active site also do not present any significant differences, apart from the hydrogen atoms that become positioned in positions that promote stronger non-bonded interaction after the MD simulation (**Fig. 28** Right).

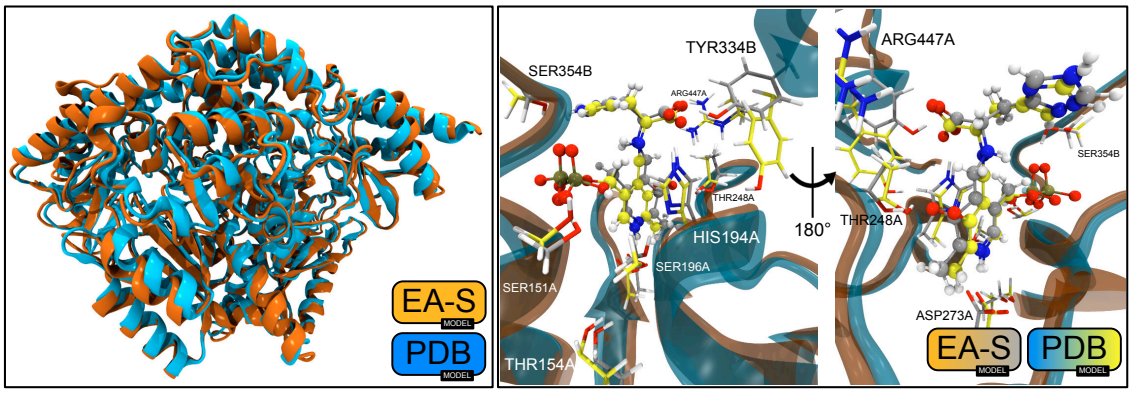



**Fig. 28 – Comparison of a representative structure from the MD simulation (EA-S model) (orange and gray colored) and the original PDB structure (blue and yellow colored).**
(Left) NewCartoon representation of the protein. (Right) CPK and licorice representations of the external aldimine and active site amino acid residues, respectively.

In the **QI** model, we found higher conformational differences between the equilibrated structure and the original PDB once the global RMSd value raises to 2.11 ± 0.11 Å (**Fig. 29** Blue line) and the backbone RMSd value reaches 1.69 ± 0.12 Å (**Fig. 29** Orange line). This difference is mostly due to the absence of the carboxylate group, which induces some conformational changes in the active site. This evidence is confirmed by a higher RMSd value (1.22 ± 0.15 Å) for the active site of the **QI** model comparatively to the **EA-S** model (**Fig. 27** Red line).

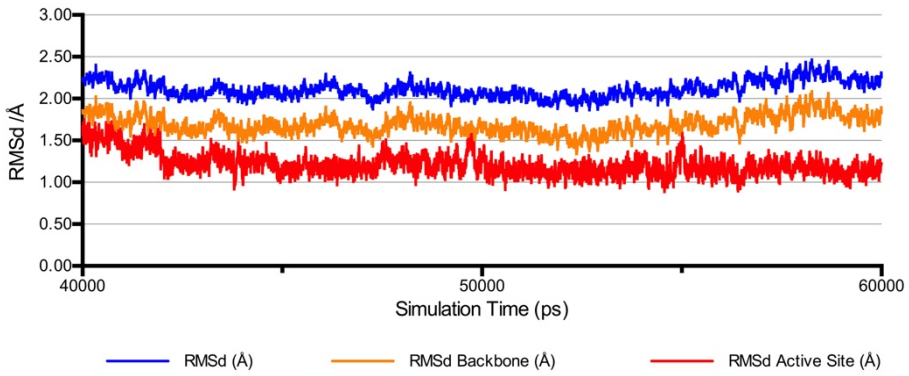



**Fig. 29 - RMSd values for the equilibrated MD simulation of the QI model using the original PDB structure as a reference.**
The blue line corresponds to the RMSd value of all atoms in the system. The orange line corresponds to the RMSd values considering only the α-carbons, corresponding to the protein backbone. The red line corresponds to the RMSd values considering the active site amino acid residues and the quinonoid intermediate.

A representative structure from the MD simulation using the **QI** model was chosen to overlap with the PDB structure. Despite the comparatively higher RMSd values, no noteworthy conformation disparities were noticed (**Fig. 30** Left). However, the amino acids residues from the active site present noteworthy differences relatively to the initial PDB structure. The most significant ones involve His-194A that no longer established a

hydrogen bond with substrate (since the carboxylate group no longer exists). In addition, and similar to what was observed with the **EA-S** model, some hydrogen bonds were also optimized, namely the ones involving the quinonoid intermediate and the amino acid residues Ser-354B, Ser-151A, Ser-196A, Thr-154A and Thr-248A (**Fig. 30** Right).

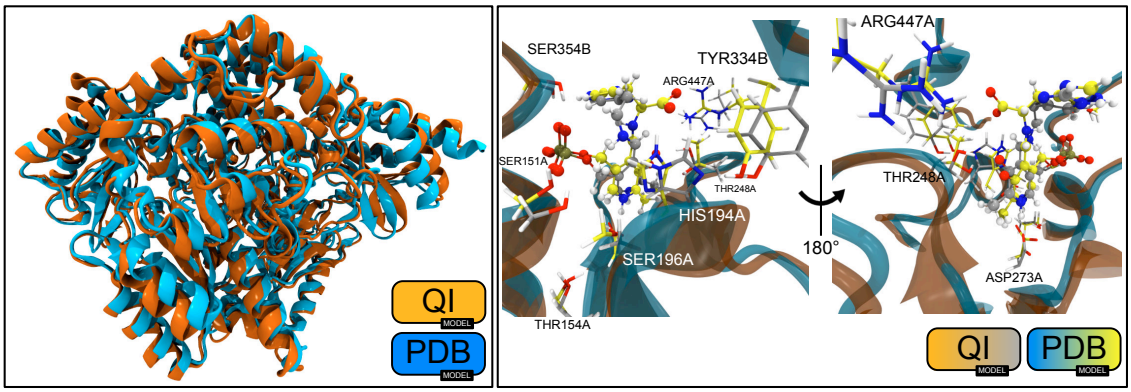



**Fig. 30 – Comparison of a representative structure from the MD simulation (QI model) (orange and gray colored) and the original PDB structure (blue and yellow colored).**
(Left) NewCartoon representation of the protein. (Right) CPK and licorice representations of the quinonoid intermediate and active site amino acid residues, respectively.

In conclusion, all the MD simulations produce equilibrated structures that do not differ too much from the initial ones but promote a better orientation of the involved atoms, and therefore a better accommodation of the external aldimine and quinonoid intermediate inside the active site of mHDC.

#### *2.2.2.2. Root-Mean-Square Fluctuations (RMSF)*

The root-mean-square fluctuation (RMSF) is a measure of the flexibility of a certain atom or amino acid residue, once it is a measure of the deviation between that particle and its averaged position during a MD simulation. We could evaluate the RMSF of different residues and understand in which extension those residues are flexible.

We analyze the RMSF values of each amino acid residue of the HL of the **EA-S** model during the MD simulation (**Fig. 31** Right). Most of the amino acid residues present RMSF values between 3 and 5 Å, but the residues closer to the active site (**Fig. 31** Right – red sections) have the lowest values. Inherently, the most flexible residues are located on the external region where the solvent exposition is greater, especially in the C-terminal region (**Fig. 31** Left).

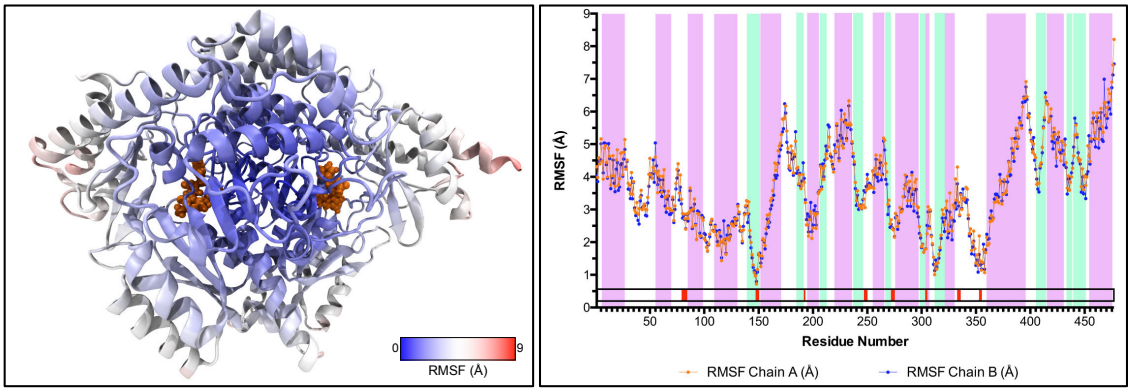



**Fig. 31 – RMSF values of each amino acid residue during the MD simulation of the EA-S model.**
(Left) NewCartoon representation of a representative structure from the MD simulation colored by RMSF value. VDW representation of the external aldimine colored by orange. (Right) RMSF values of each amino acid residue belonging to each protein subunit. Pink and green regions correspond to helices and sheets, respectively. Red regions indicate amino acid residues at the active site.

Concerning the active site residues, we found that most of them exhibit poor flexibility. This occurs due to the formation of strong hydrogen bonds between the PLP intermediates and some residues of the active site, that turn their position almost constant during the equilibrated period of MD simulation. In this regard, the stabilization of the phosphate group of the PLP cofactor by Ser-151A and Ser-354B, is particularly important, since it behaves as an anchor of the full PLP cofactor inside the active site (**Fig. 32**). In contrast, Tyr-334B is the most flexible amino acid residue (**Fig. 32**) within the active site. Its flexibility has already been observed during the RMSd analysis, once this tyrosine residue has a completely different orientation comparing to the one that it adopts in the PDB structure (**Fig. 28** Right). Such conformational flexibility of Tyr-334B was expected since it is known that it belongs to a well characterized flexible loop of the enzyme [67].

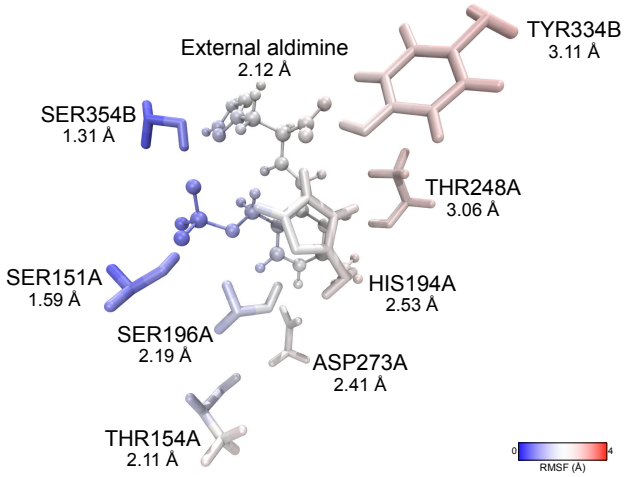



**Fig. 32 – RMSF analysis of the MD simulation using the EA-S model.**
CPK and licorice representations of the external aldimine and relevant active site amino acid residues, respectively, colored by RMSF value. The RMSF value of each residue and external aldimine are placed near each structure.

Despite the high flexibility of Tyr-334B, this residue is always in close proximity to the PLP cofactor. Taking into account that this is highly conserved among the PLP-dependent decarboxylases [29, 68], this results suggests that it can be important for catalysis [66].

Considering now the **QI** model, the RMSF evaluation revealed a similar profile (**Fig. 33** Right) relatively to the **EA-S** model. Once more, the most rigid residues are placed in the protein's core while the most flexible ones establish a direct contact with the solvent, mainly the ones from the C-terminal region (**Fig. 33** Left).

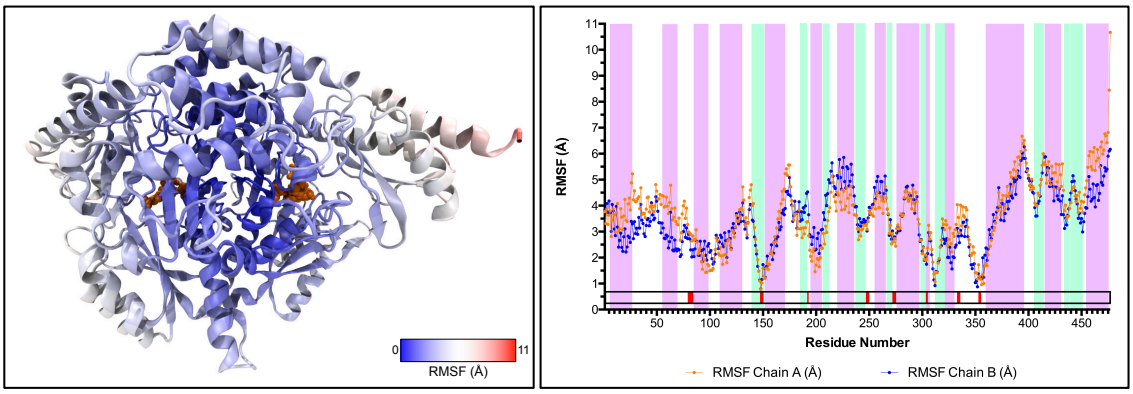



**Fig. 33 - RMSF values of each amino acid residue during the MD simulation of the QI model.**
(Left) NewCartoon representation of a representative structure from the MD simulation colored by RMSF value. VDW representation of the quinonoid intermediate colored by orange. (Right) RMSF values of each amino acid residue belonging to each protein subunit. Pink and green regions correspond to helices and sheets, respectively. Red regions indicate amino acid residues at the active site.

Looking at the active site of **QI** model, we can verify the same trend that was described for the **EA-S** model. Despite the similarities, Tyr-334B appears to be less flexible (**Fig. 34**) when the quinonoid intermediated occupies the active site, indicating a stronger interaction with the quinonoid intermediate when compared with the external aldimine of the **EA-S** model (**Fig. 32**).

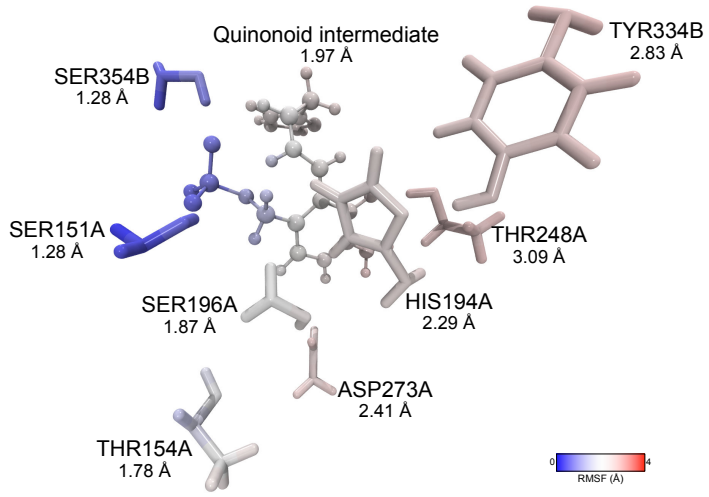



**Fig. 34 - RMSF analysis of the MD simulation using the QI model.**
CPK and licorice representations of the quinonoid intermediate and relevant active site amino acid residues, respectively, colored by RMSF value. The RMSF value of each residue and quinonoid intermediate are placed near each structure.

These results also confirm the role of phosphate as an anchor for the PLP and also show up a significant higher flexibility of the Tyr-334B, which may be important for the catalysis.

#### *2.2.2.3. Radial Distribution Function (RDF)*

In addition to the flexibility analysis, the solvent exposure of the amino acid residues is also a relevant aspect for catalysis. So, radial distribution function (RDF) can be used to measure the solvent exposure of a certain atom. Furthermore, the integral of this function gives the number of water molecules placed at a certain distance of a given atom. This type of analysis becomes interesting when we want to understand the exposure of the residues of the enzyme in relation to the solvent and disclose about the influence of the solvent in a certain chemical reaction.

When we carefully observe the active site of mHDC, three main regions are clearly more solvent exposed: the region nearby the carboxylate group of the substrate, and the regions where Tyr-334B and Asp-273A residues are located at. The RDF analysis for the carboxylate group (**Fig. 35** Top Left) and the Tyr-334B (**Fig. 35** Bottom Left) showed a high proximity to the solvent (**Fig. 35** Right). In fact, the carboxylate group has its first solvation sphere approximately at 4 Å and above that distance, the number of surrounding water molecules increases exponentially due to the solvent proximity. The solvent exposure is more evident in the case of Tyr-334B, which establishes its first solvation sphere with two water molecules, at a lower distance (3.85 Å) (**Fig. 35** Right).

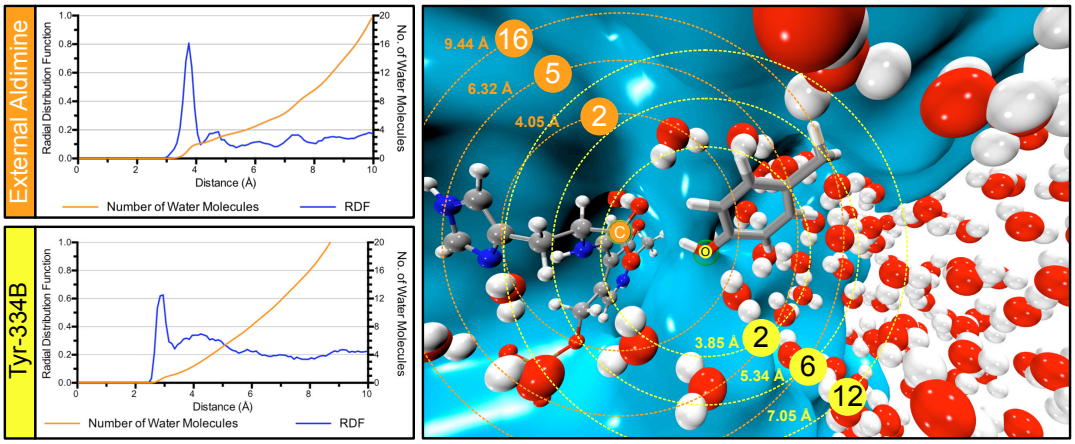



**Fig. 35 – RDF analysis of MD simulation using the EA-S model.**

(Left) RDF values as a function of the distance and the cumulative number of water molecules regarding (Top) the carboxylate group of the external aldimine and (Bottom) the hydroxyl group of the Tyr-334B. (Right) Surface and VDW representation of the subunit A and water molecules, respectively. CPK and licorice representation of the external aldimine and Tyr-334B amino acid residue, respectively. Each dashed circumference represents a different solvation sphere placed at a given distance. Colored circles indicate the cumulative number of water molecules at a certain solvation sphere.

Combining RDF and RMSF data about Tyr-334B, we could corroborate the flexibility of this tyrosine, which is associated with its position in a very flexible and solvent exposed loop. Indeed, the proteolytic susceptibility of the Tyr-338B-containing loop is justified by this high flexibility and exposure to solvent.

An important water interaction was also observed nearby Asp-273A. This aspartate residue establishes a hydrogen bond with the nitrogen N4 of the PLP and also with two water molecules (**Fig. 36** Right and Top Left). Interestingly, besides the solvent proximity of this aspartate residue, the first solvation sphere does not contact directly with the solvent (**Fig. 36** Bottom Left).

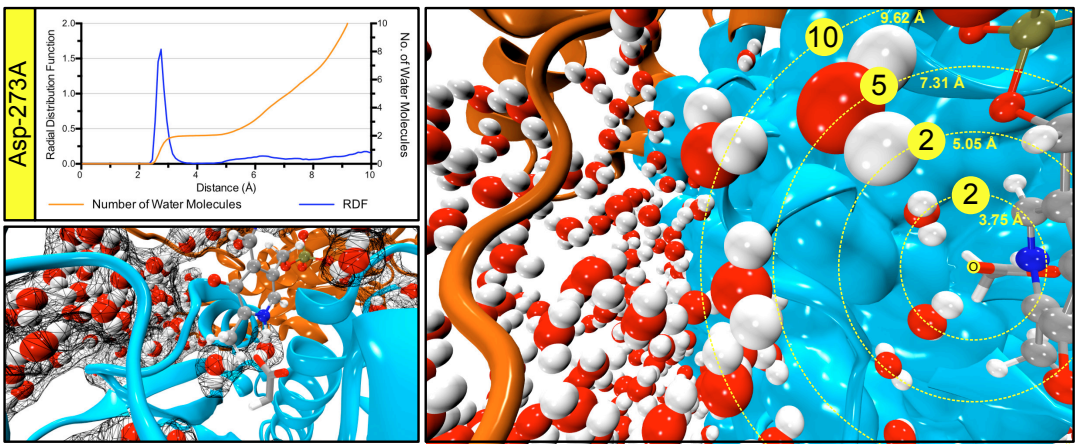



**Fig. 36 - RDF analysis of MD simulation using the EA-S model.**

(Top Left) RDF values as a function of the distance and the cumulative number of water molecules regarding the Asp-273A. (Bottom Left) Wireframe surface representation of the intermolecular interaction between the water molecules. (Right) Surface and VDW representation for the subunit A and water molecules, respectively. CPK and licorice representation for the external aldimine and Asp-273A amino acid residue, respectively. Each dashed circumference represents a different solvation sphere placed at a given distance.

These results suggested that the protein fosters the presence of water molecules nearby Asp-273A, promoting a stable and permanent interaction between those water molecules and the aspartate residue. In fact, this could explain the ubiquitous conservation of those two water molecules among all PLP-dependent decarboxylases. If we look at the PDB file of the original hHDC structure as well as to the structure of the AADC, we could confirm the presence of these two water molecules in a similar position (**Fig. 37**).

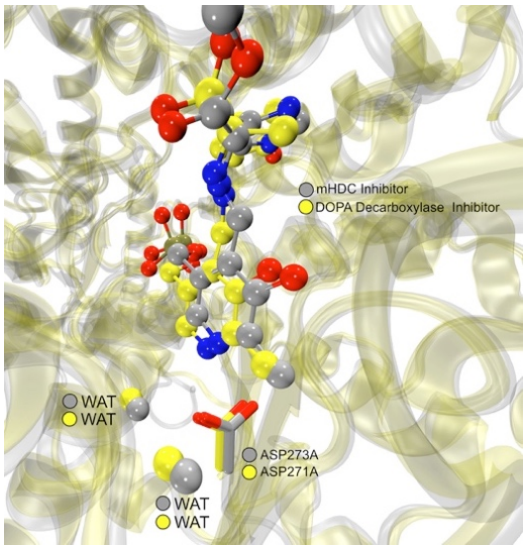



**Fig. 37 – Overlapping the PDB structures of (gray) the hHDC (PDB code: 4E1O) (gray) and (yellow) the AADC (PDB code: 1JS3) [29].**

CPK representation of the inhibitor of each enzyme and licorice representation of the aspartate residue. The conservative water molecules (oxygen atom) are represented by VDW.

Considering the **QI** model, the PLP structure remains highly exposed to the solvent, and mainly carbon Cα (**Fig. 38** Top Left). Due to the fact that the carboxylate group has been released, Tyr-334B cannot establish any intermolecular interaction with the quinonoid intermediate. However, despite its more solvent exposure, a lower number of water molecules are involved in the solvation spheres of Tyr-354B (**Fig. 38** Bottom Left and Right) comparatively with the **EA-S** model (**Fig. 35** Right).

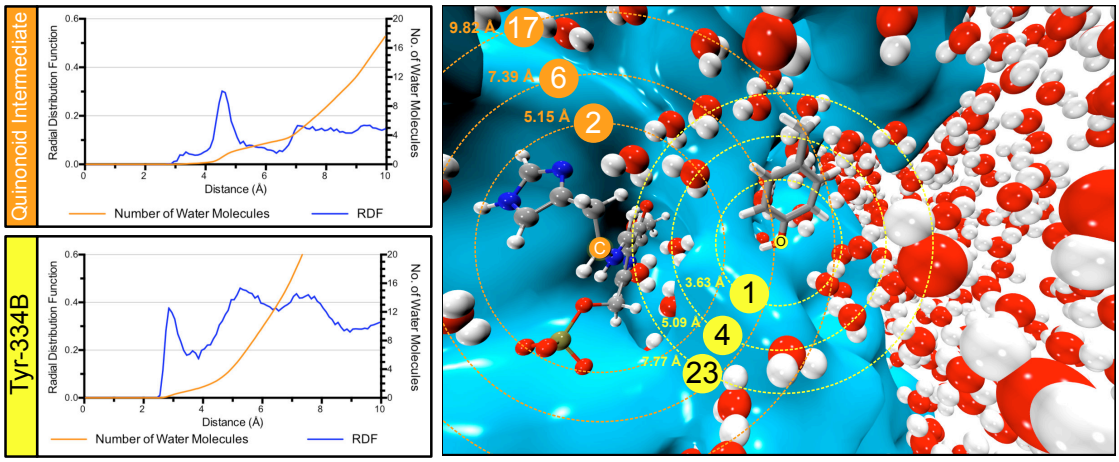



**Fig. 38 - RDF analysis of MD simulation using the QI model.**

(Left) RDF values as a function of the distance and the cumulative number of water molecules regarding (Top) the Cα of the quinonoid intermediate and (Bottom) the hydroxyl group of the Tyr-334B. (Right) Surface and VDW representation of the subunit A and water molecules, respectively. CPK and licorice representation of the quinonoid intermediate and Tyr-334B amino acid residue, respectively. Each dashed circumference represents a different solvation sphere placed at a given distance. Colored circles indicate the cumulative number of water molecules at a certain solvation sphere.

In the **QI** model, Asp-273A establishes considerable more interactions with water molecules (**Fig. 39** Right) than it was found in the **EA-S** model (**Fig. 36** Right). This evidence is supported by the formation of a new solvation sphere at 4.5 Å (**Fig. 39** Top Left) which is absent on the **EA-S** model (**Fig. 39** Top Left). In sum, Asp-273A requires a higher stabilization by the nearby water molecules, when the quinonoid intermediate is formed.

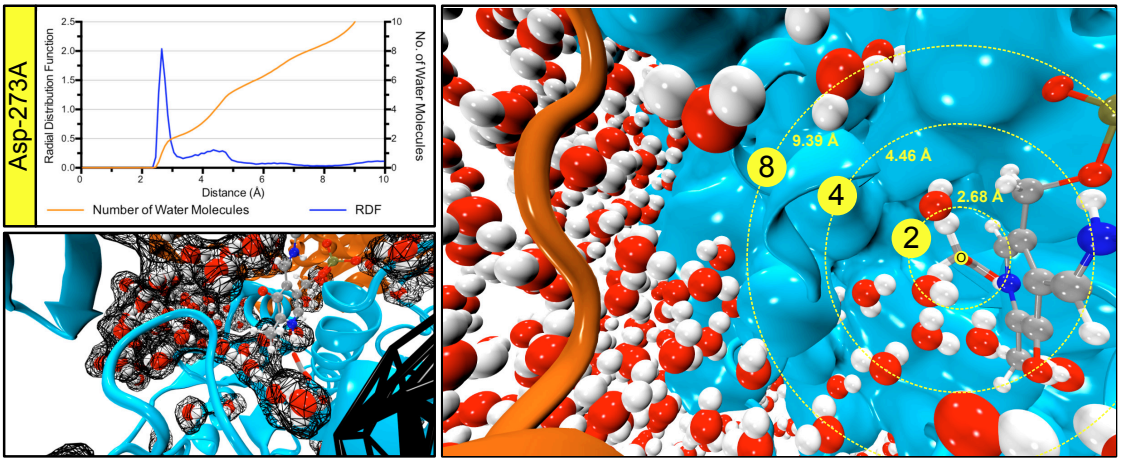



**Fig. 39 - RDF analysis of MD simulation using the QI model.**

(Top Left) RDF values as a function of the distance and the cumulative number of water molecules regarding the Asp-273A. (Bottom Left) Wireframe surface representation of the intermolecular interaction between the water molecules. (Right) Surface and VDW representation for the subunit A and water molecules, respectively. CPK and licorice representation for the quinonoid intermediate and Asp-273A amino acid residue, respectively. Each dashed circumference represents a different solvation sphere placed at a given distance.

#### *2.2.2.4. Hydrogen-bond analysis*

Non-bonded interactions are an important feature in the stabilization of biological systems, and typically they change in the active site during catalysis. Consequently, the dynamics of these interactions are crucial to understand the catalytic mechanism, in particular to unravel the critical amino acid residues that are directly involved in catalysis. In this work, a hydrogen interaction analysis along the MD simulations were done, aiming a map of the most important interactions.

Both models (**EA-S** and **QI**) present several hydrogen interactions with the phosphate group, which is understandable due its higher negative charge and electron density. In fact, five (**Fig. 40** Orange frame) or four (**Fig. 41** Orange frame) hydrogen interactions are established and maintained during the entire simulation of the **EA-S** and **QI** models, respectively.

The carboxylate group of the external aldimine (**EA-S** model) is stabilized by two hydrogen bonds promoted by Tyr-334B and His-194-A residues (**Fig. 40** Red frame). This histidine residue (His-194A) also establishes an important interaction with the Ser-196A, which helps it to become properly positioned in the active site (**Fig. 40** Red Frame). Those interactions are absent on the **QI** model, since the **QI** does not have the carboxylate group or any polar atom surrounding that position.

Nonetheless, both models have a hydrogen bond between Thr-248A and oxygen O1 of the PLP, which are preserved during the simulations and it is not disrupted after the carboxylate group release (**Fig. 40** Red frame and **Fig. 41** Red frame).

The imidazole group from the substrate slightly interacts with Pro-82A through its protonated nitrogen (**Fig. 40** Green frame). This interaction is even weaker in the **QI** model (lower that 5% of the last 20 ns of the simulation) (**Fig. 41** Green frame).

Finally, the Asp-273A residue presents a stronger interaction with the protonated nitrogen N4 in the **EA-S** model (**Fig. 40** Blue frame). However, the carboxylate group release and the consequent quinonoid intermediate formation lead to an increase in the distance between that aspartate residue and the hydrogen of the nitrogen N4. This increase is accomplished by a higher flexibility of the aspartate residue, comparatively to the **EA-S** model, culminating in a lower timely prevalence of the hydrogen bond (**Fig. 41** Blue frame).

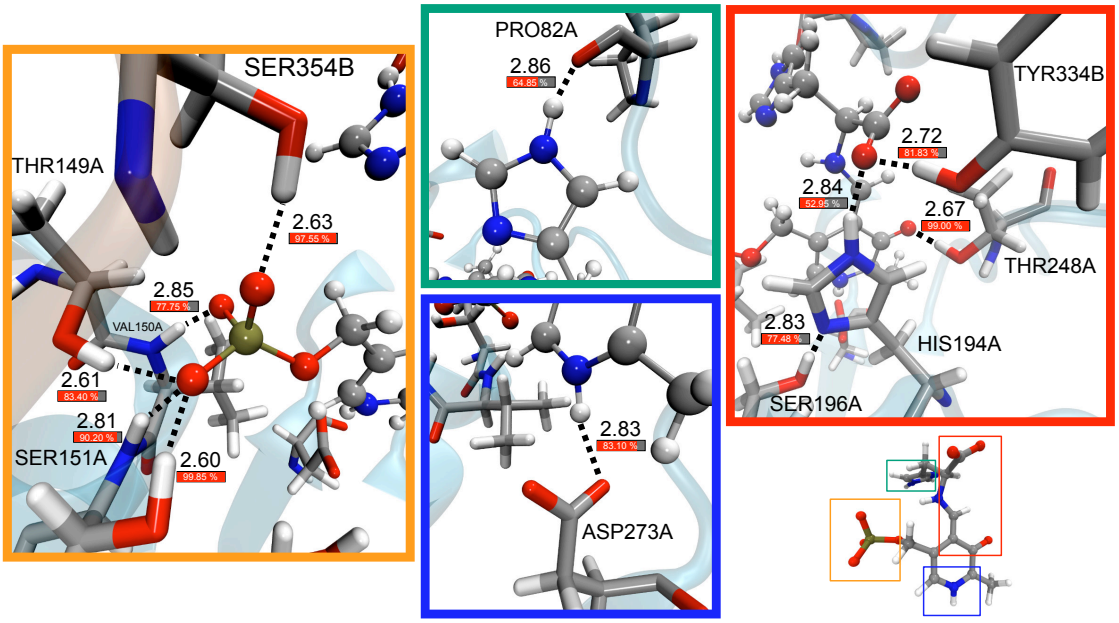



**Fig. 40 – Analysis of the hydrogen interactions between relevant active site amino acid residues (Licorice representation) and the external aldimine (CPK representation) – EA-S model.**

Dashed lines represent the hydrogen interactions, and they are accomplished by the mean distance along the MD simulation, in Å; and the respectively time prevalence.

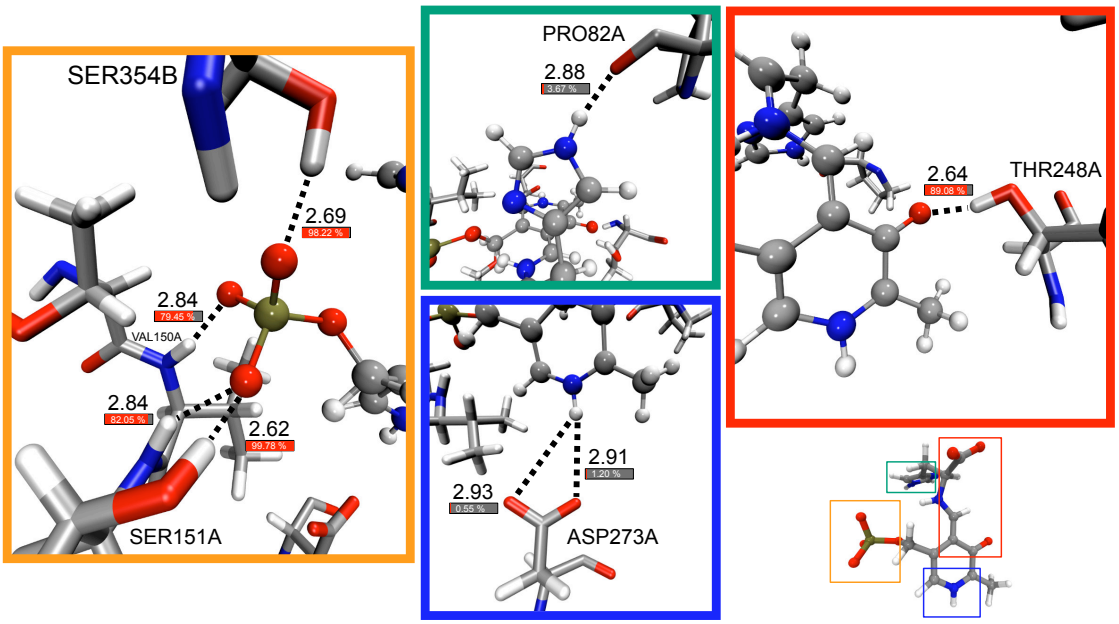



**Fig. 41 - Analysis of the hydrogen interactions between relevant active site amino acid residues (Licorice representation) and the quinonoid intermediate (CPK representation) – QI model.**

Dashed lines represent the hydrogen interactions, and they are accomplished by the mean distance along the MD simulation, in Å; and the respectively time prevalence.

This data agrees with the RDF information obtained from the MD simulations of the **QI** model that shows that the weaker interaction of Asp-273A with the PLP cofactor is extensively compensated by the stronger interactions provided by the surrounding water molecules.

All of these results gave important hints about the catalytic mechanism, and they were also essential for building the QM/MM models. Notwithstanding, we have to be careful and understand that MM information cannot be so accurately as desired and it can lead to a misinterpretation of the mechanism.

## 3. Catalytic Mechanism

The main goal of this work was the characterization of the catalytic mechanism of a mammal PLP-dependent decarboxylase from family II, called Histidine decarboxylase (mHDC), which plays a relevant role in several biological processes. We based our first trial for the catalytic mechanism of mHDC on the available experimental data [26] that proposes a two-step type of mechanism for the *L*-histidine decarboxylation. The first step is characterized by the decarboxylation of the external aldimine (EA-S) and, the consequently formation of a quinonoid intermediate (QI) and release of carbon dioxide (**Fig. 42** – Step 1). Afterward, the second step takes place and the quinonoid intermediate (QI) is protonated by an amino acid residues from the active site (**Fig. 42** – Step 2) originating a new external aldimine, but now containing histamine bonded to the PLP cofactor (EA-P). Finally, the enzymatic turnover occurs and the concomitant release of histamine (product) takes place. In this process, Lys-305A becomes once again bonded to the PLP cofactor, making the enzyme ready for a new catalytic cycle.

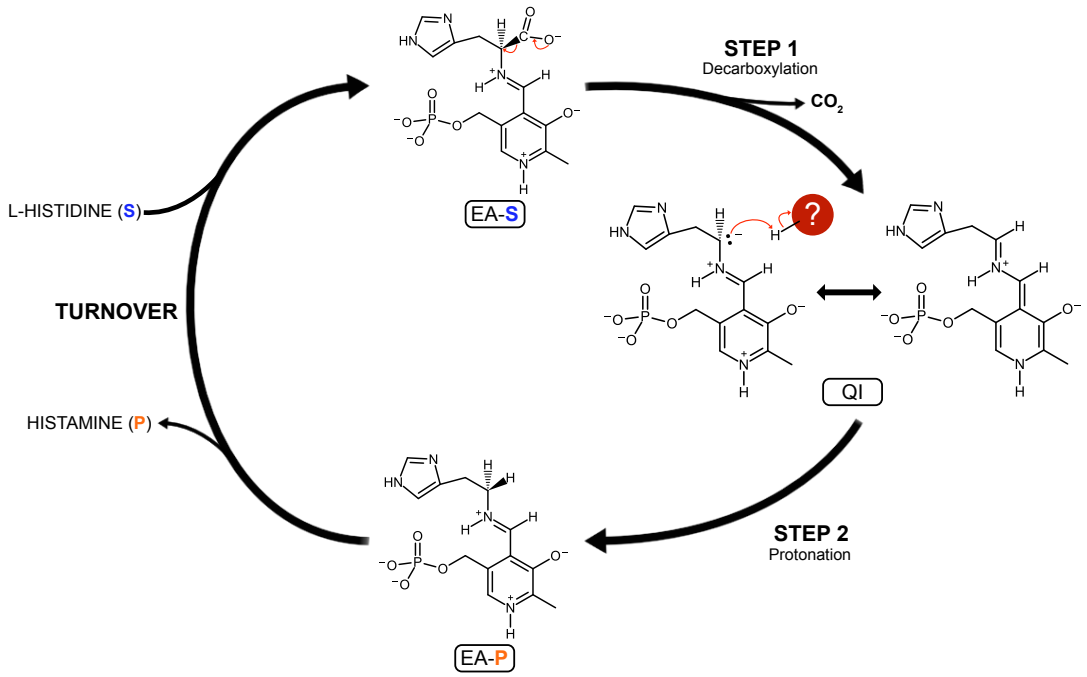



**Fig. 42 – Scheme representing the reagent, intermediates and product of the enzymatic reaction catalyzed by the mHDC.**

In order to studied these two steps of the decarboxylation process, the models retrieved from the MD simulation (**EA-S** and **QI**) were used to determine a potential energy surface (PES) through hybrid ONIOM (QM/MM) calculations performed using Gaussian09 software [178]. The most important active site residues together with the

PLP cofactor and the substrate (external aldimine or quinonoid intermediate) were included in the high-level (HL) layer, whereas the remaining part of the protein and water molecules were incorporated in the low-level (LL) layer. According to this scheme, the HL layer region was calculated using a QM method, DFT:B3LYP/6-31G(d), while the LL layer uses a MM approach considering the ff99SB and GAFF force fields [135, 173].

Since no significant rearrangements of the overall structure of the enzyme are expected during the catalytic process, all the atoms that are located more than 30 Å away from the active site were fixed during the geometry optimizations. However, all the atoms located below this threshold were maintained free during the geometry optimizations in order to accommodate all the conformational changes of the residues that are located more closely to the active site.

In all the geometry optimizations, we first searched for the transition state (TS) structure starting from a structure similar to the reactant model. This was generally obtained with uni-dimensional scans along the particular reaction coordinate in which we were interested in. Once a putative TS structure was located, and thus was fully characterized (excepted for the frozen atoms), the reactants and the products associated with it were determined after intrinsic reaction coordinate (IRC) calculations. In all cases, the geometry optimizations and the stationary points were obtained with standard Gaussian convergence criteria. The final energies of the minima and TS structures were additionally characterized by single-point energy calculations in the HL layer using the B3LYP, M06 and M06-2X functionals along with the 6-311++G(3df,2pd) basis set in the HL layer. Dispersion corrections were also included in the final energies. The LL layer was always calculated with MM methods. The full characterization of the minima of each studied reaction was achieved by the nuclear Hamiltonian calculation. The energy contribution of the nuclei is obtained through this calculation. The obtained thermal corrections allow the determination of all thermodynamic quantities ($\Delta H$, $\Delta S$, and $\Delta G$). Additionally, the nuclear Hamiltonian calculation includes a frequency analysis that it is particularly important to confirm the TS, which should have solely one imaginary frequency.

All the energy designations that will be presented in the following sections will depend on the type of calculations that were performed. The energy difference between two structures will be designated by $\Delta E$ when the energies are calculated considering only the electronic Hamiltonian. If those energies are between an optimized TS and an optimized reactant, the associate $\Delta E$ will be presented as an activation energy ($E_a$). Additionally, considering the energy difference between an optimized product and an optimized reactant, the $\Delta E$ is called reaction energy ($E_r$). Afterwards, the nuclear

Hamiltonian was calculated and the enthalpy and entropic variations ($\Delta H$ *and* $\Delta S$) could also be calculated together with the zero-point energy correction. This allowed to compute the Gibbs free energy ($\Delta G$) either for the respective $E_a$ or $E_r$.

In this chapter, the terms exothermic and endothermic will be used when the $E_r$ is negative or positive, respectively. Otherwise, if we consider the $\Delta G_r$, the adopted designation will be exergonic or endergonic for negative or positive values of $\Delta G_r$, respectively.

The manipulation and analysis of the Gaussian files (input and output ones) were done using the GaussView 5.0.8. [179].

### 3.1. Step 1 – Decarboxylation

The first step of the decarboxylation process was studied using an averaged structure from the most representative cluster (among 5 clusters) obtained from the MD simulation of the **EA-S** model. This structure was then used to build the QM/MM models. During the study of the first step of the decarboxylation process, several different QM/MM schemes of this model were used. In all of these schemes, the same reaction coordinate was followed. However, in each one of them, a different number of atoms was included in the HL layer and therefore in the LL layer of the QM/MM scheme. The choice of the atoms that were included in the HL layer was done in an incremental mode. This means that in the first scheme, only a few sets of amino acid residues were included in the HL layer together with the external aldimine containing the PLP cofactor bonded to *L*-histidine. In the next schemes, the number of amino acid residues included in the HL layer increased. The choice of these amino acids was always done interactively based on visual inspection, but also on single-point energy calculations that help us to choose which amino acid residues could stabilize the charge delocalization of the reaction intermediates along the catalysis (that was found to be an important aspect of the mechanism).

As the results were obtained, it was also required to make some structural modification of the model system that was retrieved from the MD simulations. These new models contain different conformation rearrangement or displacement of some amino acid residues or in the structure of the external aldimine. In the majority of these models a different QM/MM schemes were also used.

In order to prevent any misconception along the thesis, in the following sections, the QM/MM models that were built will be called as **EA-S-Mx** or **QI-Mx** according to the step that we are describing: the decarboxylation of the external aldimine or the

protonation of the quinonoid intermediate, respectively. The "**x**" is the number of the model, and the models with the number 1 correspond to the ones which were directly obtained from the MD simulations (**EA-S-M1**, **QI-M1A** and **QI-M1B**). The subsequent models are modified versions of these models, e.g. increase of the size of the HL layer or structural modifications in the structure of the external aldimine or some active site residues. These models got the numbers 2, 3 and 4. When small modifications were done, and it requires a side-by-side comparison with the model that gave rise to it, the " ' " symbol was added at the end of the designation (example: **EA-S-M1** and **EA-S-M1'**). In other cases, a same model was used to test two different reaction coordinates, therefore the letters **A** and **B** were added at the end of the designation of each model (example: **QI-M1A** and **QI-M1B**).

### 3.1.1. Model EA-S-M1: Tyr-334B displacement

The first QM/MM model (**EA-S-M1**) came from the MD simulation as mentioned before. In this model, 96 atoms were included in the high level (HL) layer belonging to the **EA-S**, together with the side chains of the following amino acid residues: Asp-273A, His-194A, Ser-151A, Tyr-334B, Ser-354B, Thr-248A and Ser-196A (**Fig. 43**).

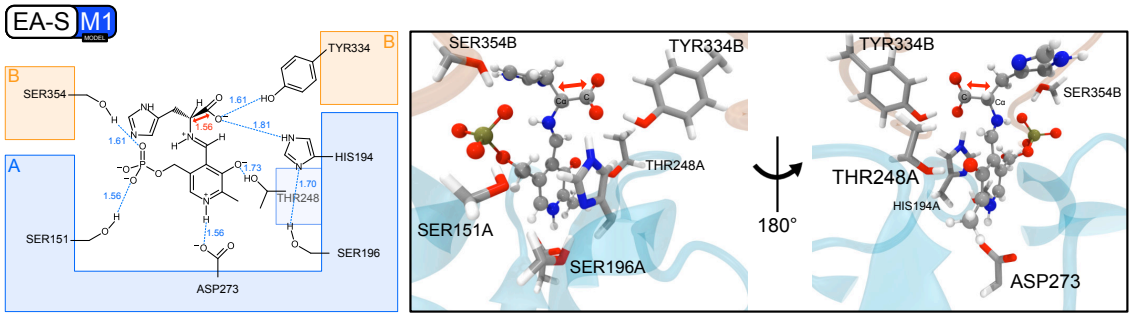



**Fig. 43 – Structure of the EA-S-M1 model.**
(Left) Wedge-Dash representation of the external aldimine (reactant) and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased during the scan, and respectively interatomic distance in Å. (Right) CPK and licorice representations of the external aldimine and HL amino acid residues, respectively.

The entire system was optimized using the QM/MM scheme: B3LYP/6-31G(d):AMBER(ff99SB+GAFF). We first optimized the structure of this model and then started the search for the TS structure increasing the bond length between the Cα-C atoms through a linear scan, that promote the decarboxylation process. In spite of several attempts have been conducted, it was not observed a decrease in the energy of the scan. The energy profile of this reaction is also very unfavorable. The energy difference between the structure where the interatomic distance between Cα-C atoms

was 2.40 Å and the reactant was 41.94 kcal/mol (**Fig. 44** Top), indicating that such reaction could not be possible under physiological conditions. A closer inspection of the model indicated that the reason for the high energies observed in this trial were due to the close proximity of Tyr-334B to the carboxylate group that prevents the release of carbon dioxide (**Fig. 44** Bottom). Consequently, the decarboxylation becomes more difficult and even impossible.

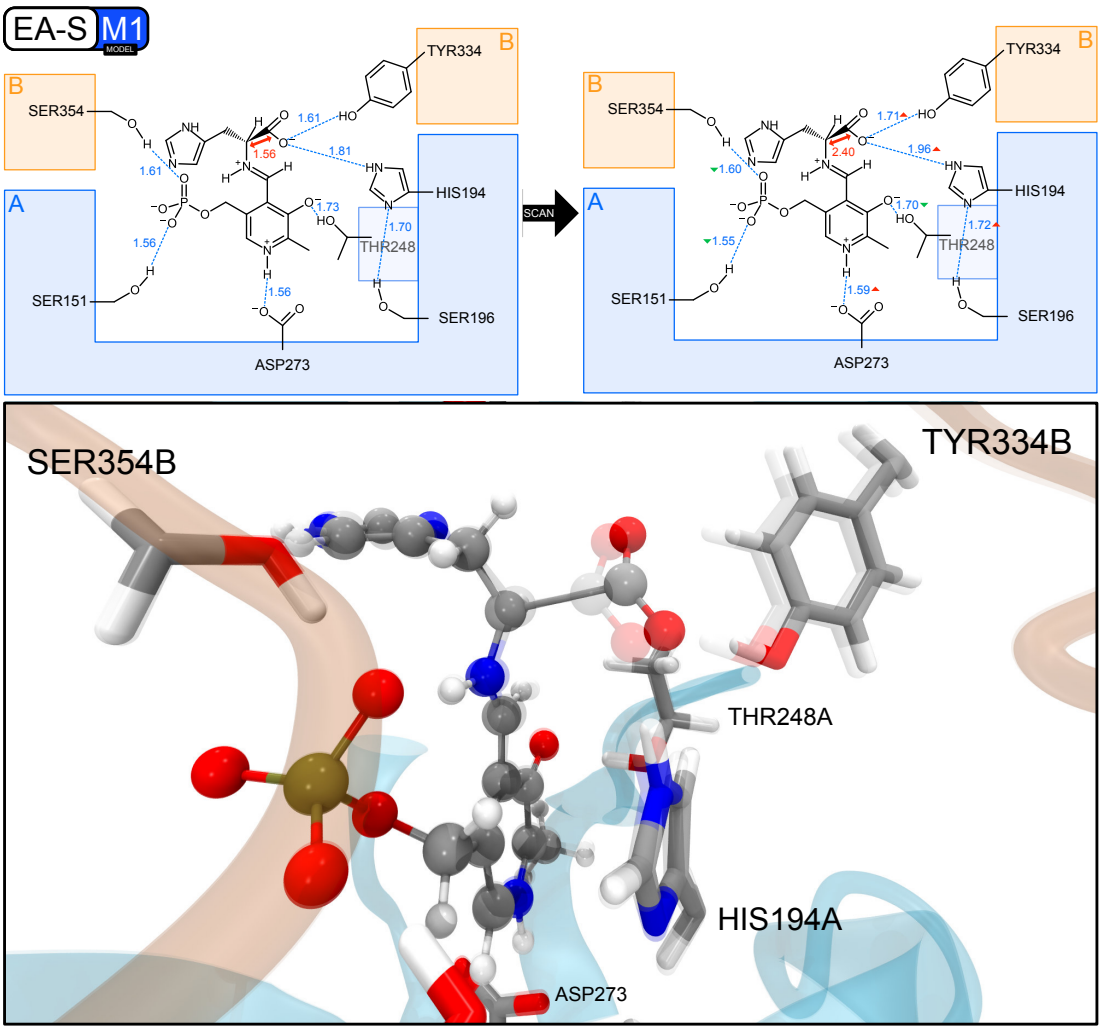



**Fig. 44 – Comparison between reactant and last structure from scan.**
(Top) Wedge-Dash representation of the external aldimine (reactant) and HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. (Bottom) Structural comparison between the reactant (lower opacity) and the last structure from the scan (higher opacity). CPK and licorice representations of external aldimine and HL amino acid residues, respectively.

The RMSF analysis of MD simulation for the Tyr-334B residue (**Fig. 32**) has already indicated that this residues is extremely flexible. This means that this residue is able to move away from the active site towards the solvent. In fact, this flexibility is promoted by its water exposure, which is also corroborated by the RDF profile (**Fig. 35**

Bottom Left). The MD simulation visualization also confirmed this data, since Tyr-334B can be highly or less exposed to the solvent whenever it is in close contact to the solvent or establishes a hydrogen bond with carboxylate group of **EA-S**, respectively.

Bearing in mind these observations, a new model (**EA-S-M1'**) was built through a slight displacement between Tyr-334B and the carboxylate group of **EA-S** (**Fig. 45**). Again, the structure was optimized using the same QM/MM scheme and the optimized complex was used to perform a linear scan where the Cα-C bond length was increased.

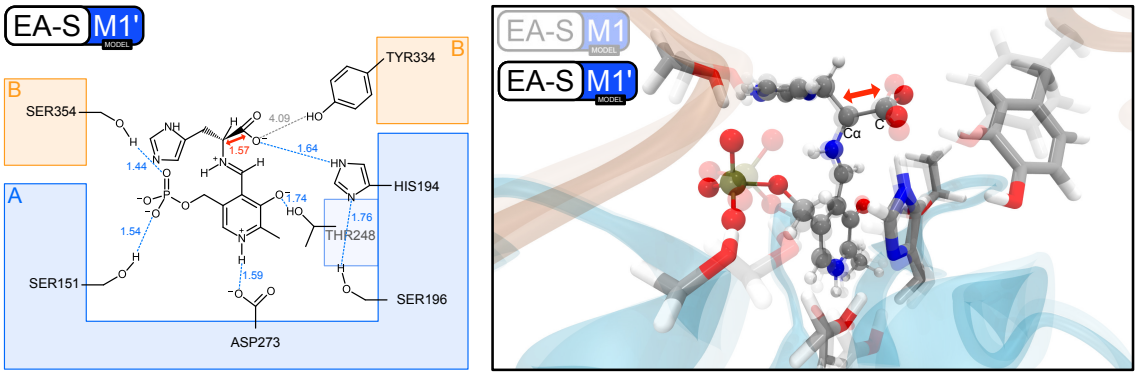



**Fig. 45 - Structure of the EA-S-M1' model.**
(Left) Wedge-Dash representation of the external aldimine (reactant) and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. Gray dashed line represents the distance between Tyr-334B and the carboxylate group of the external aldimine. (Right) Comparison between models **EA-S-M1** (lower opacity) and **EA-S-M1'** (higher opacity). CPK and licorice representations of the external aldimine and HL amino acid residues, respectively.

With this new model and considering the same interatomic distance between Cα-C atoms (2.40 Å), the computed energy for the decarboxylation process was 30.02 kcal/mol (**Fig. 46**), which is significant lower than the one that was obtained in the previous model (**EA-S-M1**). However, and despite the energy decrease in the energy required for the decarboxylation process, no TS was assessed, indicating, once again, that such reaction is not feasible.

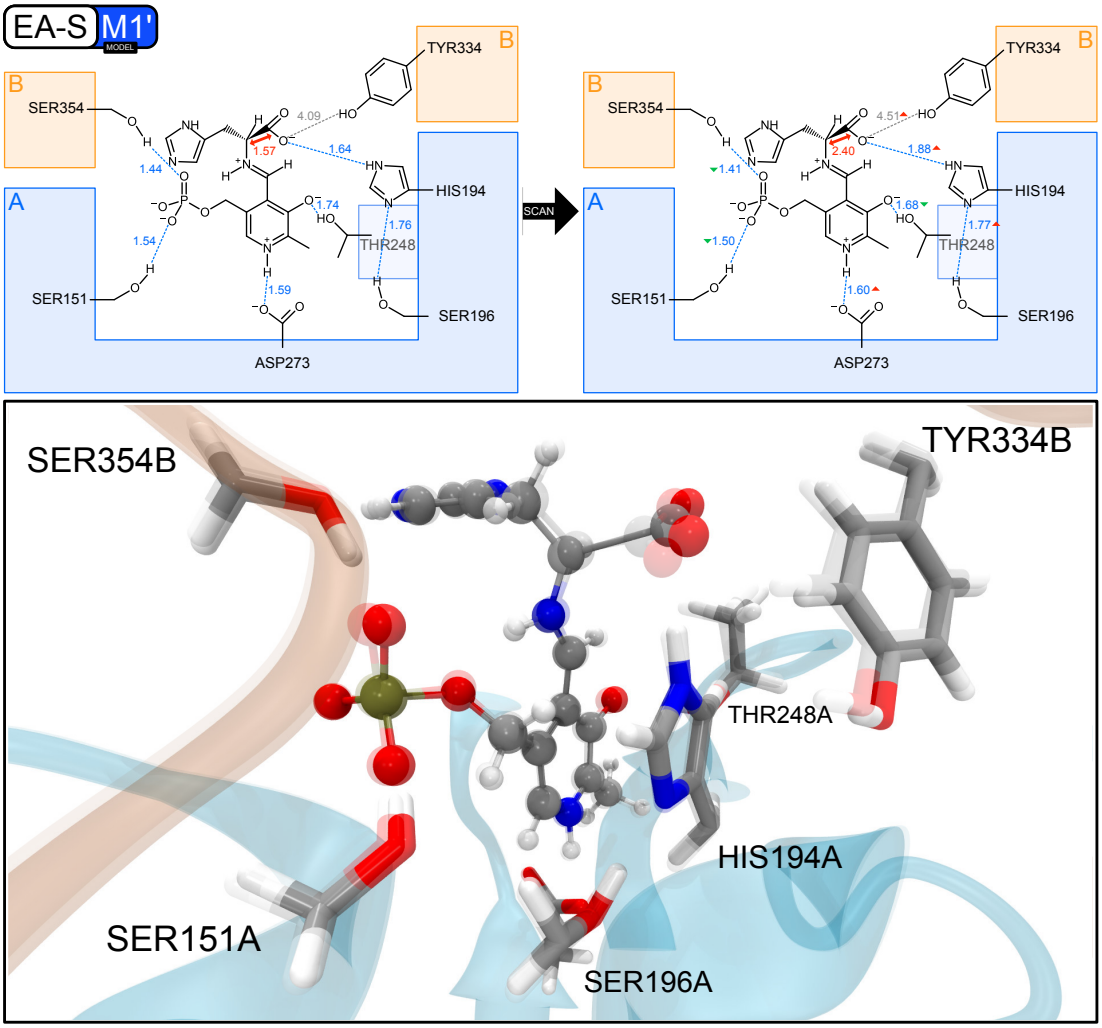



**Fig. 46 - Comparison between reactant and last structure from scan.**
(Top) Wedge-Dash representation of the external aldimine (reactant) and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. Gray dashed line represents the distance between Tyr-334B and the carboxylate group. (Bottom) Structural comparison between the reactant (lower opacity) and the last structure from the scan (higher opacity). CPK and licorice representations of the external aldimine and HL amino acid residues, respectively.

A new model (**EA-S-M1''**) was then built increasing substantially the distance between Tyr-334B and the carboxylate group of external aldimine comparatively to the previous model (**EA-S-M1'**) (**Fig. 47**).

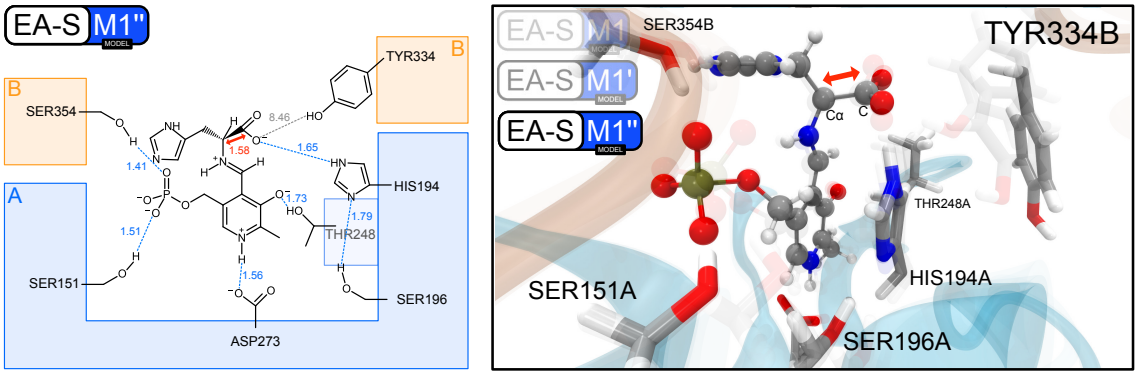



**Fig. 47 - Structure of the EA-S-M1'' model.**

(Left) Wedge-Dash representation of the external aldimine (reactant) and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. Gray dashed line represents the actual distance between Tyr-334B and the carboxylate group. (Right) Comparison between models **EA-S-M1** (lower opacity), **EA-S-M1'** (intermediate opacity) and **EA-S-M1''** (higher opacity). CPK and licorice representations of the external aldimine and HL amino acid residues, respectively.

The scan was repeated using the same QM/MM scheme and increasing the interatomic distance Cα-C. However, the energy profile of the reaction decreases only by ~5 kcal/mol comparing with model **EA-S-M1'** (decreases to 25.22 kcal/mol) (**Fig. 48**). Unfortunately, no TS was obtained once again, confirming an impossible reaction under the current conformation.

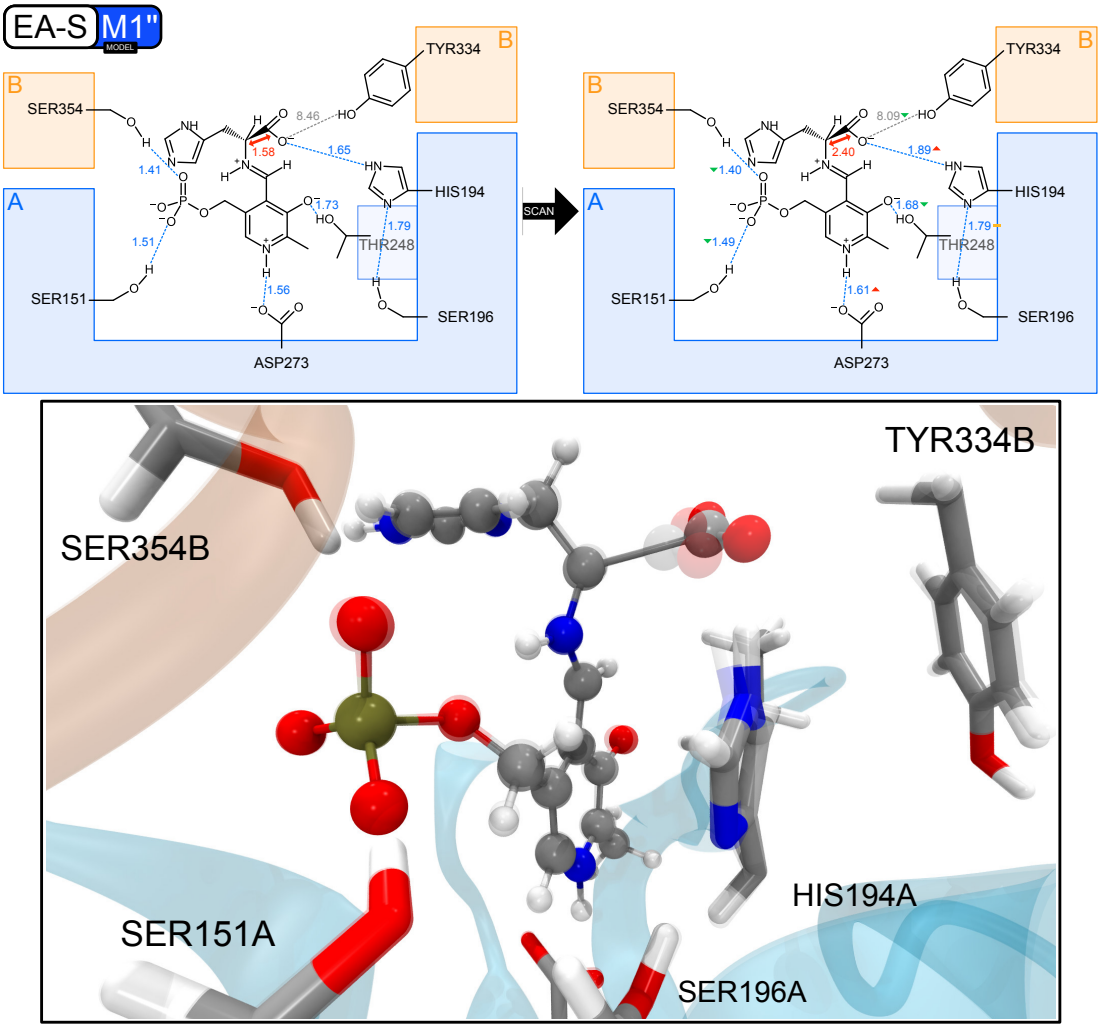



**Fig. 48 - Comparison between reactant and last structure from scan.**
(Top) Wedge-Dash representation of the external aldimine (reactant) and HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. Gray dashed line represents the actual distance between Tyr-334B and the carboxylate group. (Bottom) Structural comparison between the reactant (lower opacity) and the last structure from the scan (higher opacity). CPK and licorice representations of the external aldimine and HL amino acid residues, respectively.

These results indicate that the displacement of the Tyr-334B from the carboxylate group is an important aspect for the catalysis, since the increase of the distance between Tyr-334B and the carboxylate group from the external aldimine leads to a decrease of $\Delta E$ (**Fig. 49**). Nevertheless, this displacement is not enough to make the reaction possible, once no TS could be characterized under this structural arrangement.

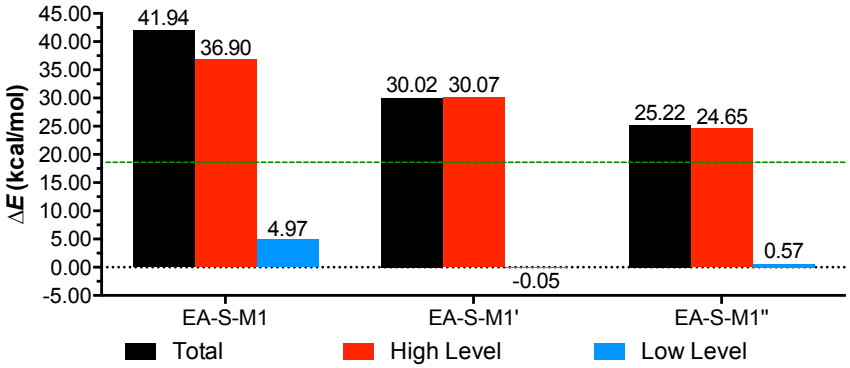



**Fig. 49 – Energy variation (Δ*E*) for each model EA-A-M1, EA-A-M1' and EA-AM1''.**
Black, red and blue bars correspond to the energy of all the system, HL layer and LL layer, respectively. The energy difference was calculated between the optimized reactant and a structure from the scan where the distance between the Cα-C was 2.40 Å. The green dashed line represents the $E_a$ predicted by the experimental $k_{cat}$.

### 3.1.2. Model EA-S-M2: Improvement of the hydrogen bond interactions

The analysis of the RDFs and hydrogen bond interactions from the MD simulation using the **EA-S** model revealed the presence of two important water molecules close to Asp-273A. The same water molecules are also present in the original structure of the PDB file, indicating that their presence might be important for catalysis. One of them is placed between Asp-273A and Ser-196A and closes a network of hydrogen bonds that also include His-194A and the external aldimine (**Fig. 50**). This result suggests that this water molecule might be important to stabilize, among the active site residues, the charge that is formed once the decarboxylation takes place.

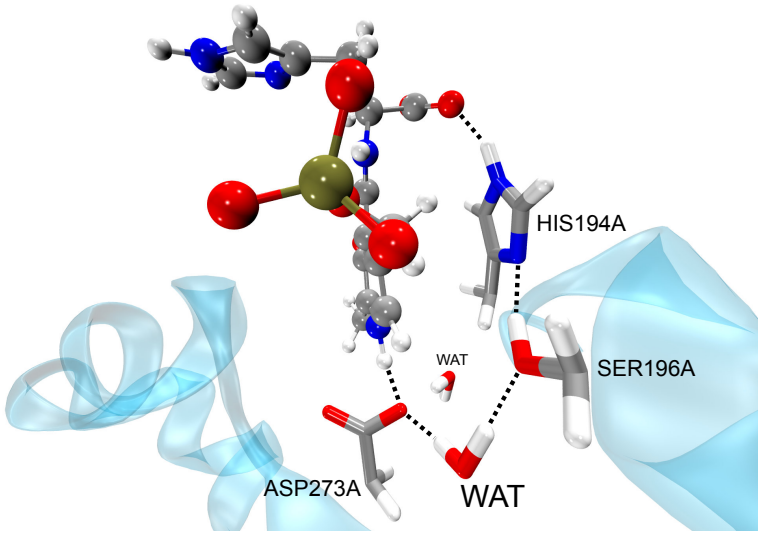



**Fig. 50 - Hydrogen interactions between the carboxylate group and Asp-273A.**
Dashed lines represent hydrogen interactions. CPK and licorice representation of the external aldimine and HL amino acid residues.

Based on this previous analysis, we built a new model (**EA-S-M2** model), contemplating this water molecule in the HL layer, using the model **EA-S-M1''** as starting point, where the Tyr-334B is displaced from its original location. However, in this model, Tyr-334B was not considered in the HL layer because it is far away from the external aldimine and should not participate on the reaction, taking into account the results obtained with previous models. In this new **EA-S-M2** model, we have also included Thr-154A in the HL layer due to its close interaction with the water molecule. Then, the system was optimized, and the resulting structure (**Fig. 51**) was used as the new reactant in which the Cα-C bond length was increased likewise in the previous calculations, through a linear scan.

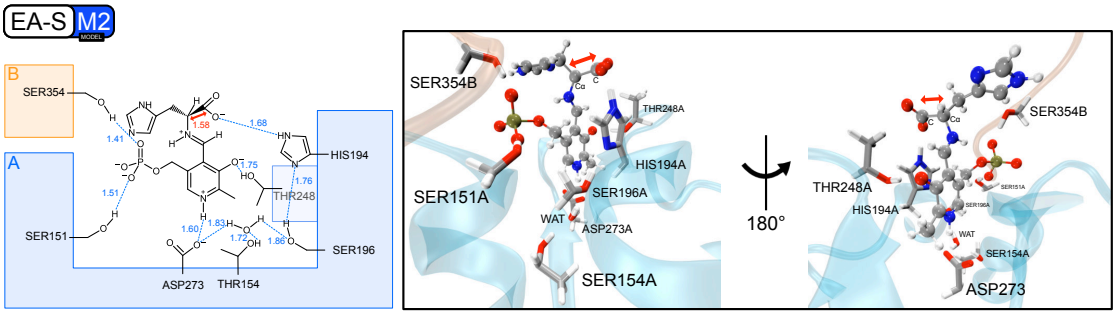



**Fig. 51 - Structure of the EA-A-M2 model.**
(Left) Wedge-Dash representation of the external aldimine (reactant), the HL amino acid residues and the structural water molecule. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. CPK and licorice representations of the external aldimine, HL amino acid residues and water molecule, respectively.

This model allowed us, for the first time, to obtain a TS that was confirmed by a solely imaginary frequency at 101.0160i $cm^{-1}$ (**Fig. 52**). The calculated activation energy was however too high ($E_a$ = 28.85 kcal/mol) (**Fig. 52**) which exceeds considerably the predictable and suitable energetic barrier. Comparing with the experimental value ($E_a$ = 18 kcal/mol), the calculated $E_a$ is too elevated [17]. This would lead to an increase in the half-life time of catalysis from 578 ms (based on experimental $k_{cat}$) to more than 2.5 years, making it impossible to occur under physiological conditions. Additionally, the product is also marginally stabilized relatively to the TS, leading to a very endothermic reaction ($E_r$ = +28.44 kcal/mol) (**Fig. 52**).

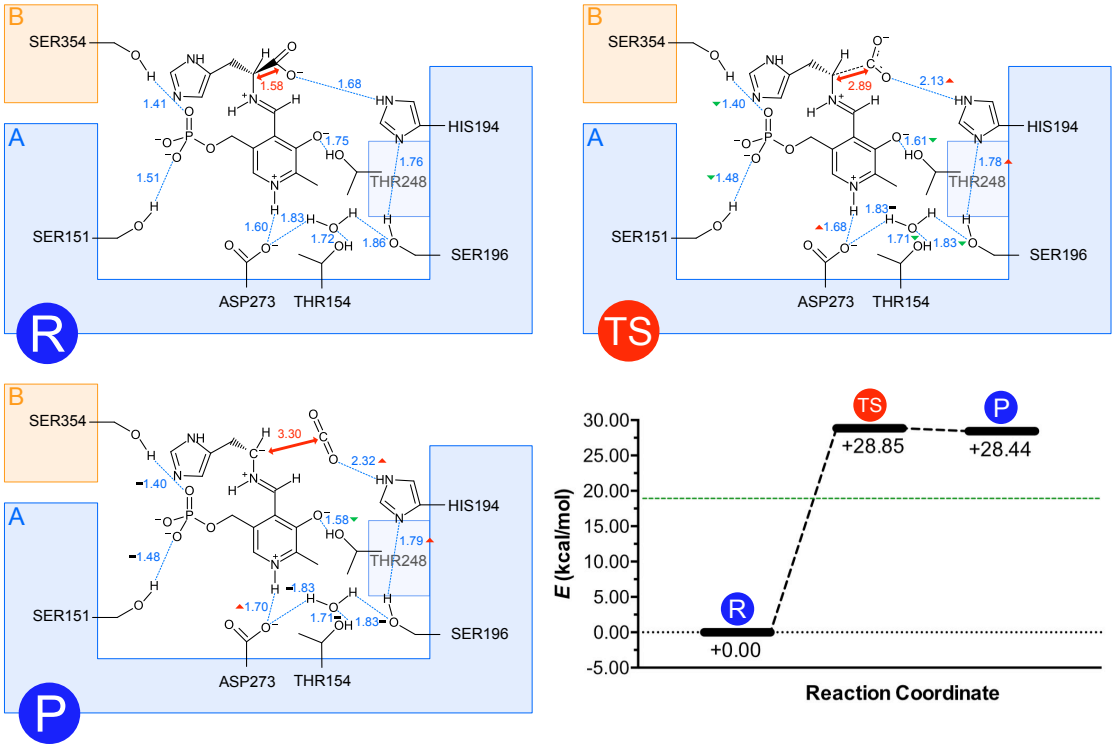



**Fig. 52 – Structure of reactant (P), transition state (TS) and product (P), and energy profile for the decarboxylation step.**

(Top and Bottom Left) Wedge-Dash representation of the external aldimine and HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. (Bottom Right) Energy profile comparing reactant, transition state and product. The green dashed line represents the energy barrier predicted by $k_{cat}$.

A closer look to the $E_a$ and $E_r$ considering the contribution of each QM/MM layer, revealed a major contribution of the HL layer. In other words, the energy variation between TS and reactant (**Fig. 53**), as well as, the difference between product and reactant (**Fig. 53**) are mainly affected by the atoms considered under the HL layer. Consequently, the high activation and reaction energies are due directly to the atoms involved in the HL model. The contribution of the LL layer for these energies is very small and below 1 kcal/mol. These results indicate that perhaps more amino acids from the active site need to be included in the HL layer or that these energies reflect some problems related with the position and orientation of the external aldimine and/or surrounding amino acid residues.

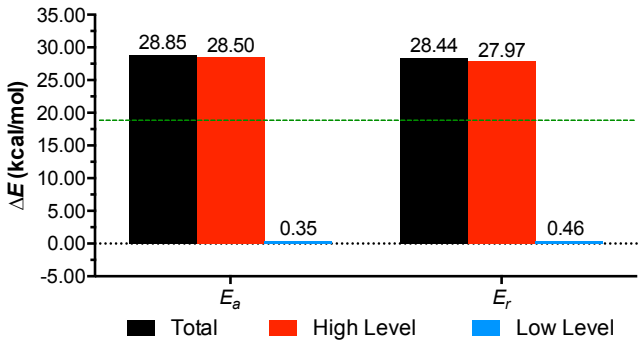



**Fig. 53 – Activation ($E_a$) and reaction ($E_r$) energies.**
Energies values considering all atoms of the system system (black), the HL layer (red) or the LL layer (blue). The green dashed line represents the activation energy predicted by the $k_{cat}$.

A closer look to the relevant interactions between the external aldimine and the amino acid residues of active site revealed a weakening of the interaction between His-194A and the carboxylate group of the external aldimine in the TS structure (**Fig. 52**). This indicates that the carboxylate group of external aldimine became less negative, which turns the reaction less favorable (**Table 3**). In the same way, the hydrogen bond between the proton bonded to nitrogen N4 and Asp-273A also becomes weaker, since the distance increases 0.08 Å from the reactant to the TS (reactant: 1.60 Å vs TS: 1.68 Å) (**Fig. 52**). This suggests that during the reaction there is a possible charge distribution along the PLP leading to a less positive charge on nitrogen N4, culminating into a less strong interaction with Asp-273A. The charge decrease at Asp-273A (**Table 3**) justifies this previous observation, once a weaker interaction between these residues promotes a higher charge concentration on the aspartate residue.

**Table 3 - Mulliken charges (atomic units) assigned to external aldimine, carboxylate group/carbon dioxide and Asp-273A for each reactional state (reactant, transition state and product).**

| Structure | Charge (a.u.) | | |
|---|---|---|---|
| | Reactant (R) | Transition State (TS) | Product (P) |
| External Aldimine | -1.737490 | -1,733655 | -1,733655 |
| Carboxylate Group/Carbon Dioxide | -0.082569 | -0,052407 | -0,052407 |
| Asp-273A | -0.764877 | -0,766671 | -0,766671 |

Based on these results, we conclude that the present model does not represents accurately the active site during the mHDC catalysis, once the activation energies exceeds the predicted experimental value and are not acceptable under physiological conditions.

**Fig. 54** shows a comparison of the energy profiles of the models presented before. Only in the last model (**EA-S-M2** model), it was possible to obtain a TS albeit the high activation and reaction energies that were obtained.

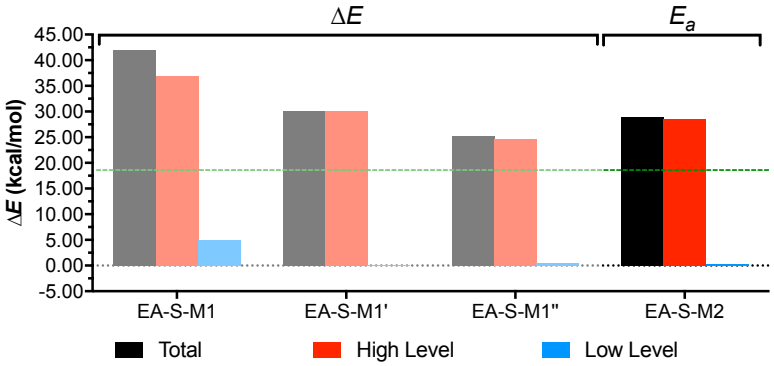



**Fig. 54 – Energy profile comparison between models.**
Black, red and blue bars correspond to the energy of all the system, HL layer and LL layer, respectively. The green dashed line represents the $E_a$ predicted by the experimental $k_{cat}$.

### 3.1.3. Model EA-S-M3: Imidazole group protonation

A closer look at the imidazole group of *L*-histidine (substrate) revealed a lack of any interaction between its protonated nitrogen (Nε) and an amino acid residue within the active site. To overcome this situation, we rotate the imidazole group by 180° and changed the nitrogen that was protonated. Consequently, the resulting model (**EA-A-M3**) enabled a strong hydrogen interaction between the protonated nitrogen (Nη) and the Ser-354B (**Fig. 55**).

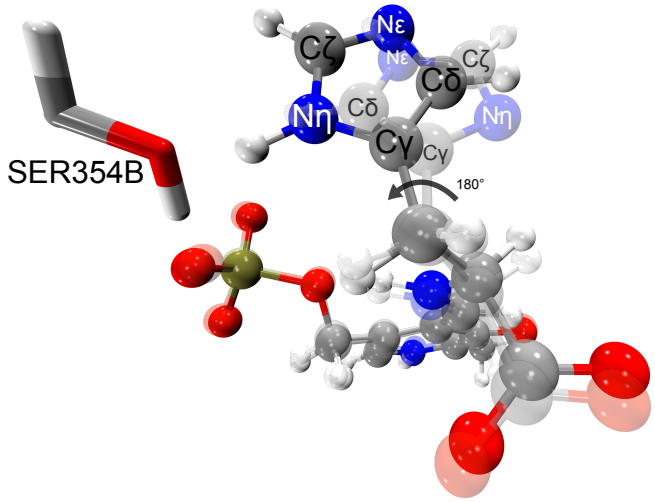



**Fig. 55 - Hydrogen interaction between the imidazole group from substrate and Ser-354B after its rotation and the change of the protonated nitrogen atoms.**

This new model (**EA-A-M3**) was built through the modification of previous model (**EA-A-M2**) and subsequent optimization. The resulting structure (**Fig. 56**) was then used to search for the reactant, TS and product of the reaction. The search for the TS was once again pursuit through linear scans involving the increase of the C-Cα bond length.

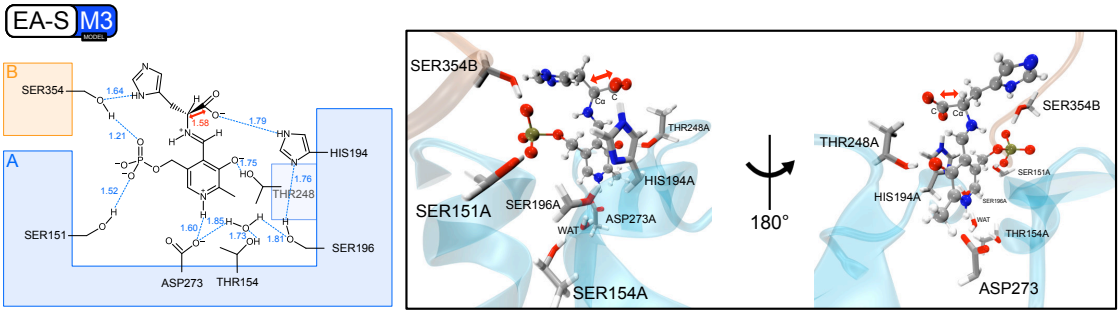



**Fig. 56 - Structure of the model EA-A-M3.**

(Left) Wedge-Dash representation of the external aldimine (reactant), the HL amino acid residues and a water molecule. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. (Right) CPK and licorice representations of the external aldimine; HL amino acid residues and water molecule, respectively.

The TS was assessed and confirmed by a unique imaginary frequency at 90.6370i $cm^{-1}$ (**Fig. 57**). The reactant and the product of the reaction were optimized from the optimized TS resulting in an activation energy of 29.19 kcal/mol (**Fig. 57**) which is indeed very similar to the one that was obtained with the previous model (**EA-S-M2**). However, if we consider only the energetic barrier associated to the HL model, 22.55 kcal/mol (**Fig. 58**), the energy is significant lower that the one obtained using the **EA-A-M2** model (**Fig. 53**). It means that the HL layer of the **EA-A-M3** model describes more accurately the active site during the decarboxylation step than the **EA-A-M2** model.

However, the activation and reaction barriers are being severely penalized in about 6 kcal/mol by the enzyme (LL layer) (**Fig. 58**). This means that somehow the enzyme is not helping to catalyze the reaction. In spite of this negative trend, this model confirmed the observations done considering the **EA-S-M2** model, where the distance between the nitrogen N4 and Asp-273A increased during the reaction (**Fig. 57**). This conformational change is a consequence of an electronic delocalization along the PLP cofactor, decreasing the hydrogen donor effect of nitrogen N4. At the same time, the water molecule placed nearby Asp-273A became closer to this residue, helping to stabilize the negative charge (**Table 4** and **Fig. 57**). Considering an overall view of entire active site, the intermolecular interactions tend to become stronger to stabilize the negative charge introduced after the carbon dioxide release (**Fig. 57**).

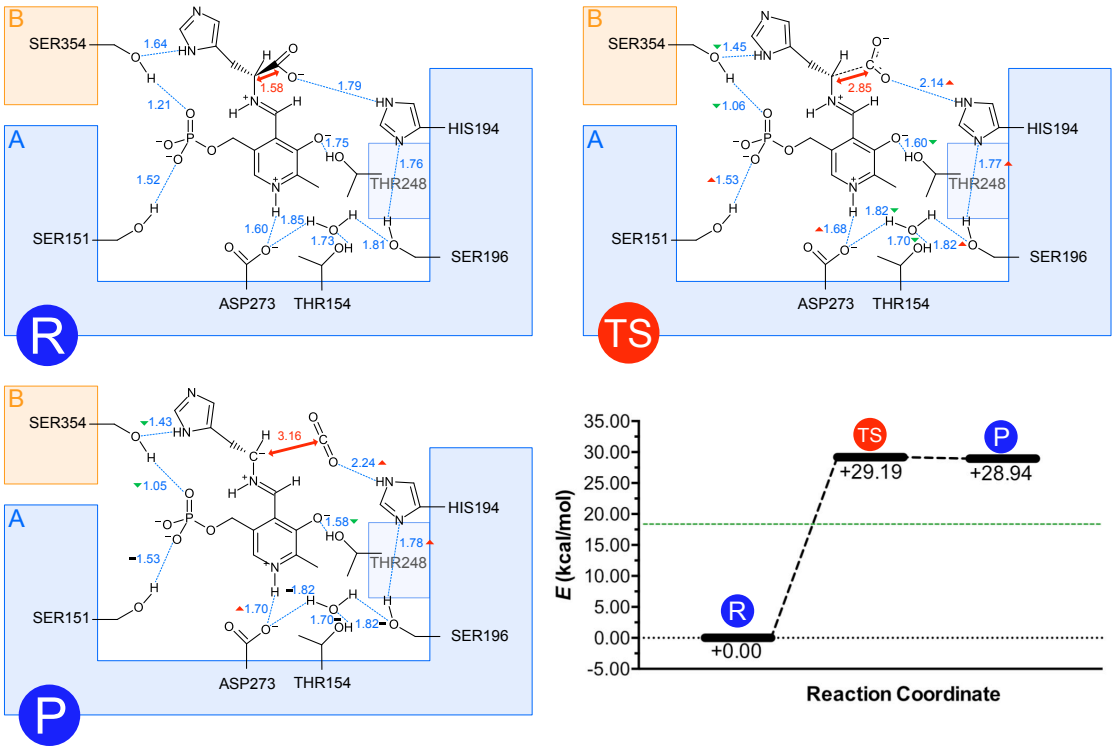



**Fig. 57 - Reactant (R), transition state (TS) and product (P) structures and energy profile.**

(Top and Bottom Left) Wedge-Dash representation of the external aldimine, structural water molecule and the HL amino acid residues. Blue dashed lines represent relevant non bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. (Bottom Right) Energy profile comparing reactant, transition state and product. The green dashed line represents the activation energy predicted by the turnover number ($k_{cat}$).

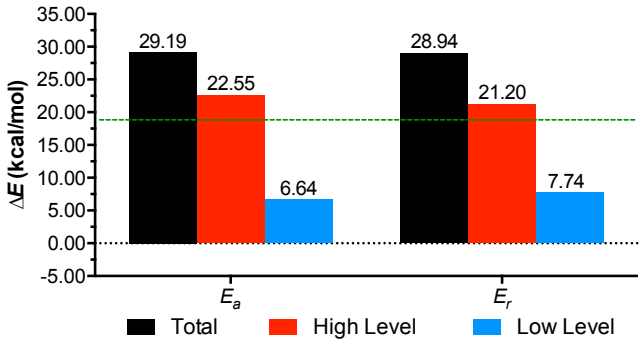



**Fig. 58 - Activation ($E_a$) and reaction ($E_r$) energies.**

Energies values considering all atoms of the system system (black), the HL atoms (red) or the LL atoms (blue). The green dashed line represents the $E_a$ predicted by the $k_{cat}$.

**Table 4 - Mulliken charges (atomic units) assigned to external aldimine, carboxylate group/carbon dioxide and Asp-273A for each reactional state (reactant, transition state and product).**

| Structure | Charge (a.u.) | | |
|---|---|---|---|
| | Reactant (R) | Transition State (TS) | Product (P) |
| External Aldimine | -1.712884 | -1.705630 | -1.697700 |
| Carboxylate Group/Carbon Dioxide | -0.110189 | -0.074334 | -0.023766 |
| Asp-273A | -0.762634 | -0.762357 | -0.765171 |

Comparing the energies obtained with this new model and the previous ones, we can conclude that the modification endorsed to the initial structure, that was retrieved from the MD simulation, has improved the description of the active site and external aldimine to promote catalysis (**Fig. 59**). In this last model, a great improvement of the energies in the HL layer have been observed and the total energy is only penalized by the unfavorable energies that are obtained by the LL layer.

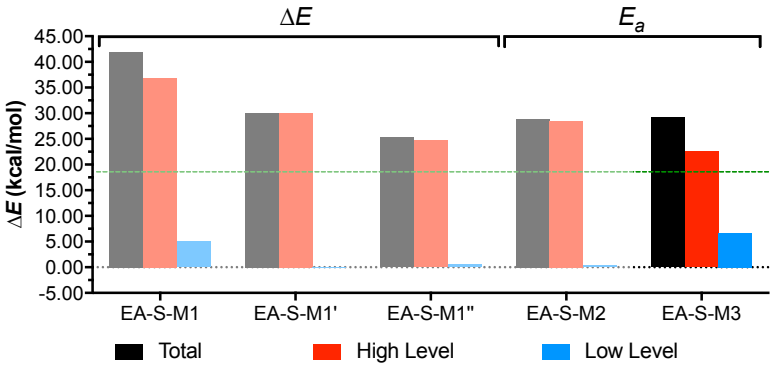



**Fig. 59 - Energy profile comparison between models.**
Black, red and blue bars correspond to the energy of all the system, HL layer and LL layer, respectively. The green dashed line represents the $E_a$ predicted by the experimental $k_{cat}$.

### 3.1.4. Model EA-S-M4: N-H bond rotation

We should emphasize, at this point, that the conformation of the imine linkage used in the previous models is what is observed in the PDB structure that contains an inhibitor bonded to the PLP cofactor. However, such conformation is not common in the PLP-dependent enzymes that have the N-H pointing towards the opposite direction.

The results of the previous models indicate the imine linkage in such orientation is not efficient for catalysis under physiological conditions, and therefore it can be an outcome of the inhibition process. In order to prove this idea, we have built a new model (**EA-S-M4**), in which the N-H bond orientation of the imine linkage of the external aldimine was changed. The re-orientation of the NH group was done through the opposite rotation, by 180°, of the two adjacent dihedrals in order to point the N-H bond to O1 oxygen (**Fig. 60**).

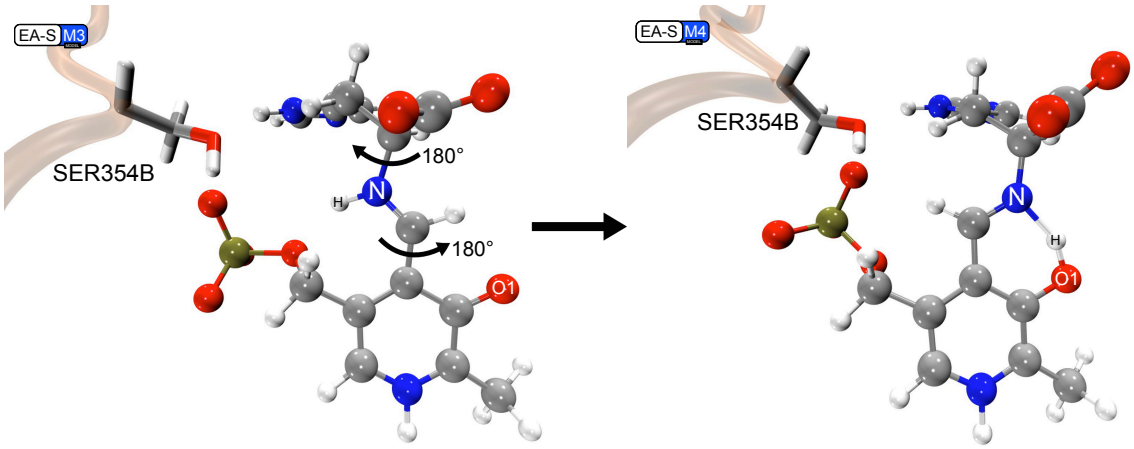



**Fig. 60 – N-H bond rotation through the rotation of its adjacent dihedral angles in opposite directions.** The model **EA-A-M4** was obtained from the model **EA-A-M3** according to the described rotations and subsequently optimization.

Regarding this new model and analyzing their suitable resonance structures (**Fig. 61**), we understand that the **EA-S-M4** model originates a quinonoid intermediate with less punctual charges. The previous model (**EA-S-M3**) presents two possible resonance structures for reactant, but when the carboxylate group is released, the resulting quinonoid intermediate presents two more punctual charges than the model **EA-S-M4**. This evidence shows up a lower capability of the **EA-S-M3** model to spread the charge along PLP, whereas when the N-H bond is rotated in the **EA-S-M4** model, the formed quinonoid intermediate does not have any punctual charge in addition to the phosphate ones.

Since the active site environment is predominantly hydrophobic, avoiding the formation of punctual charges leads to a better stabilization of the full complex. Therefore, the new model (**EA-S-M4**) comprises a better configuration of the full system that should favor the decarboxylation reaction. However, this requires a previous reaction that promotes the proton transfer from nitrogen N to oxygen O1. We have done this calculation and show that this reaction has a very low activation barrier (0.26 kcal/mol), which is also confirmed by the spontaneous proton passage during the optimization procedure.

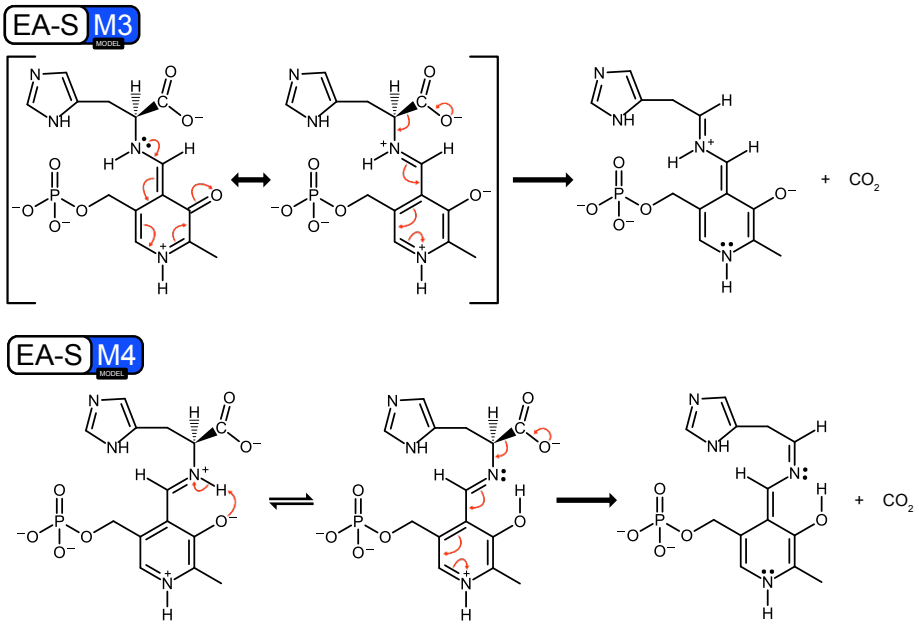



**Fig. 61 – PLP structures along the decarboxylation step considering models EA-S-M3 or EA-S-M4.**
Wedge-dash representation of external aldimine, respective resonance hybrid and intermediates along the first step.

The new model **EA-A-M4** was then optimized and using the same approach used in the previous sections. The energy profile of the reaction was assessed through a linear scan where the bond length C-Cα was progressively increased (**Fig. 62**).

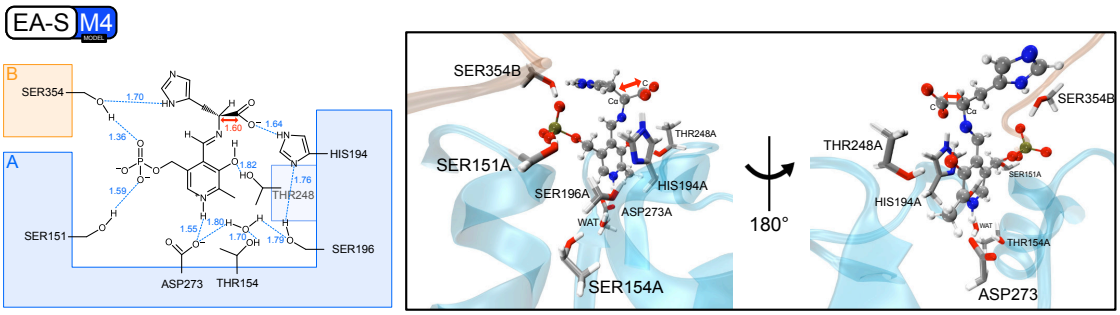



**Fig. 62 - Structure of the model EA-A-M4.**
(Left) Wedge-Dash representation of the external aldimine (reactant), the HL amino acid residues and a water molecule. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. (Right) CPK and licorice representations of the external aldimine; HL amino acid residues and water molecule, respectively.

The energy profile of the linear scan revealed that the energy required to obtain the TS decreases substantially when compared with the previous model (23.70 kcal/mol vs 29.19 kcal/mol). This barrier has a significant contribution from the LL layer, once those atoms are responsible by an energy penalization of 5.84 kcal/mol, while the HL layer contributes only by 17.86 kcal/mol for the activation barrier (**Fig. 64**). This penalization was not expected since the enzyme should promote the stabilization of the TS and therefore have favorable contribution during catalysis. Consequently, these

results indicate that the energies of this step are still too high when compared to the experimental value. Based on this fact, instead of characterizing the reactant, TS and product for this reaction we tried to search for the motives for this energetic profile in this new model.

We then realized that this penalization could be caused by the non-inclusion of certain amino acid residues in the HL layer, which could be essential to stabilize the external aldimine and promote the reaction. To provide some insight about this matter, a single-point energy (SP) calculation were done with an extended HL layer where more amino acid residues were included in the HL layer and considered under the high theoretical level. The new model (**EA-S-M4'**) includes some amino acid residues from the active site that surround specifically the phosphate group of PLP cofactor (**Fig. 63**).

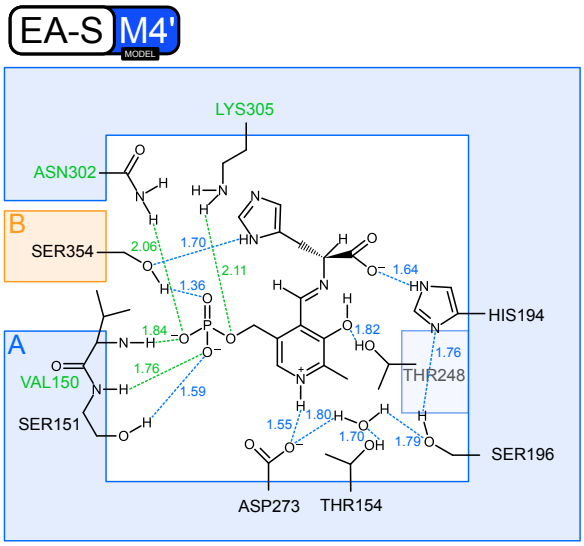



**Fig. 63 – Structure of EA-S-M4' model.**
Wedge-Dash representation of the external aldimine (reactant), the HL amino acid residues and a water molecule. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Green dashed lines represent new non-bonded interactions promoted by adding amino acid residues (green colored) to the HL layer comparatively to the **EA-S-M4** model.

We should note that these calculations did not include the optimization of the atoms of the new QM/MM scheme. The structures were just retrieved from the linear scan, one resembling the TS (the highest energy structure) and two other for the adjacent minima, one for the reactants and another for the product of the reaction. The results of these calculations with the **EA-S-M4'** model revealed a significant decrease in the $E_a$ (**Fig. 64**). This indicated a higher stabilization of the TS when the phosphate group of PLP is surrounded by more amino acid residues. These calculations also revealed that the enzyme is also able to decrease the activation barrier by 3.41 kcal/mol. Notwithstanding, the energy associated with the HL layer increased comparatively to the

**EA-S-M4** model, but this increase can be overcome with the optimization of the system. Therefore, **EA-S-M4'** model was optimized using the same QM/MM scheme (B3LYP/6-31G(d):AMBER[GAFF+ff99SB]) and the resultant reactant, TS and product of the reaction characterized (**Fig. 65**).

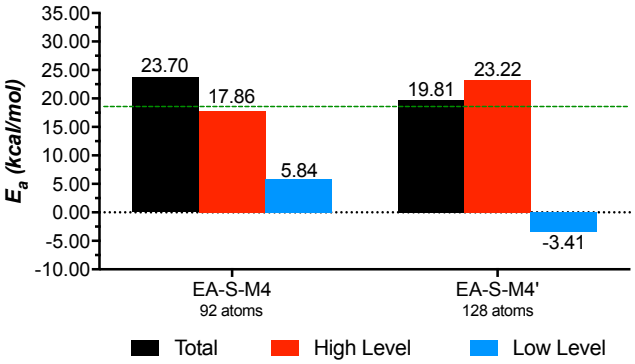



**Fig. 64 – Activation energies ($E_a$) assigned to each model.**
Energies values considering all atoms of the system system (black), the HL atoms (red) or the LL atoms (blue). The green dashed line represents the $E_a$ predicted by the $k_{cat}$.

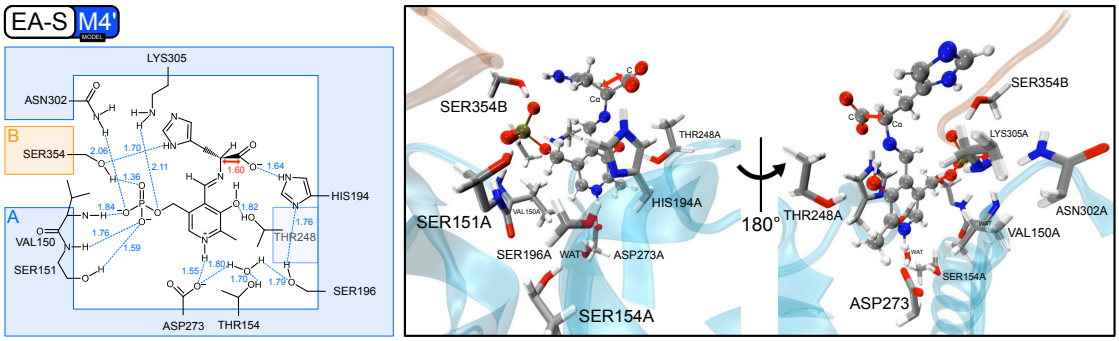



**Fig. 65 - Structure of the model EA-A-M4'.**
(Left) Wedge-Dash representation of the external aldimine (reactant), the HL amino acid residues and a water molecule. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. (Right) CPK and licorice representations of the external aldimine; HL amino acid residues and water molecule, respectively.

The **EA-S-M4'** model confirmed that an extensive stabilization of the phosphate group decreases the activation barrier, once the $E_a$ decreased from 23.70 kcal/mol to 17.12 kcal/mol. This energy has, as predictable, a major contribution of the HL atoms, 17.47 kcal/mol, while the remaining enzyme (LL layer) slightly decreases the activation barrier by 0.35 kcal/mol. As observed with previous models, we confirmed a highly endothermic step since the $E_r$ is 16.90 kcal/mol. (**Fig. 66**)

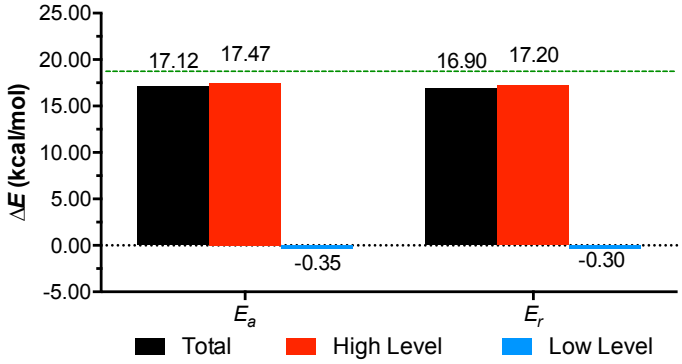



**Fig. 66 - Activation ($E_a$) and reaction ($E_r$) energies.**
Energies values considering all atoms of the system system (black), the HL atoms (red) or the LL atoms (blue). The green dashed line represents the $E_a$ predicted by the $k_{cat}$.

The TS of this reaction is characterized by an imaginary frequency at 87.2080i $cm^{-1}$. A careful analysis of the reaction reveals that the break of the bond between Cα-C atoms is triggered by the electronic density withdrawing mediated by the PLP cofactor. The oxygen O'2 from substrate and the protonated nitrogen of His-194A establish a hydrogen bond that orientates the carboxylate group of substrate during the decarboxylation. During the decarboxylation reaction, this interaction becomes weaker once the negative charge of the carboxylate group is transferred to PLP (**Fig. 68**ab) and the carbon dioxide molecule becomes non-polar (**Fig. 68**c) (interaction length increases from 1.63 Å to 2.18 Å (**Fig. 67**)). The released carbon dioxide molecule presents a charge of -0.06 a.u., which is practically neutral in contrast with the original charge of carboxylate group, -0.62 a.u. (**Fig. 68**c and **Table 5**). Consequently, the negative charge is accommodated by the PLP ring (**Fig. 68**a), since its global charge varied from +0.73 to +0.34 and it is dispersed along all ring (**Table 5**). Notwithstanding, the nitrogen N4 presents the highest charge variation (reactant: -0.35 a.u. vs product: -0.53 a.u.) indicating a major role in the PLP capability to withdraw the electronic density from Cα-C bond. This data is corroborated by a decrease in the bong length of N4-H (reactant: 1.10 Å vs product: 1.06 Å (**Fig. 67**)) and a consequently distancing from Asp-273A (reactant: 1.52 Å vs product: 1.62 A (**Fig. 67**)). Therefore, the negative charge of Asp-273A became less stabilized and the intermolecular interaction with the closer water molecule became stronger (interaction length decreases from 1.81 Å to 1.77 Å (**Fig. 67**)). Subsequently, all the hydrogen bonds established with Ser-196A and His-194A becomes weaker once all intermolecular interactions distances became higher.

With respect to the upper region of PLP (**Fig. 68**a), we can disclose the formation of double bonds between Cα-N and C6-C7 atoms, since their length decreased from 1.45 Å and 1.48 Å to 1.32 Å and 1.41 Å, respectively (**Fig. 67**). Simultaneously, the double bond between C8-N atoms was converted into a single bond (bond length

increased from 1.29 Å to 1.36 Å (**Fig. 67**)). These results indicate an electronic delocalization along those atoms but the negative charge formed due to the decarboxylation reaction is not accommodated here once the charge for that region did not varies significantly between the reactant and the product of the reaction (charge increases 0.06 a.u.) (**Table 5**).

Concerning about the interactions between the amino acids close to the phosphate group of PLP cofactor, no significant differences were observed during the reaction, indicating its major role on anchoring PLP during the catalytic process. (**Fig. 67** and **Fig. 68**)

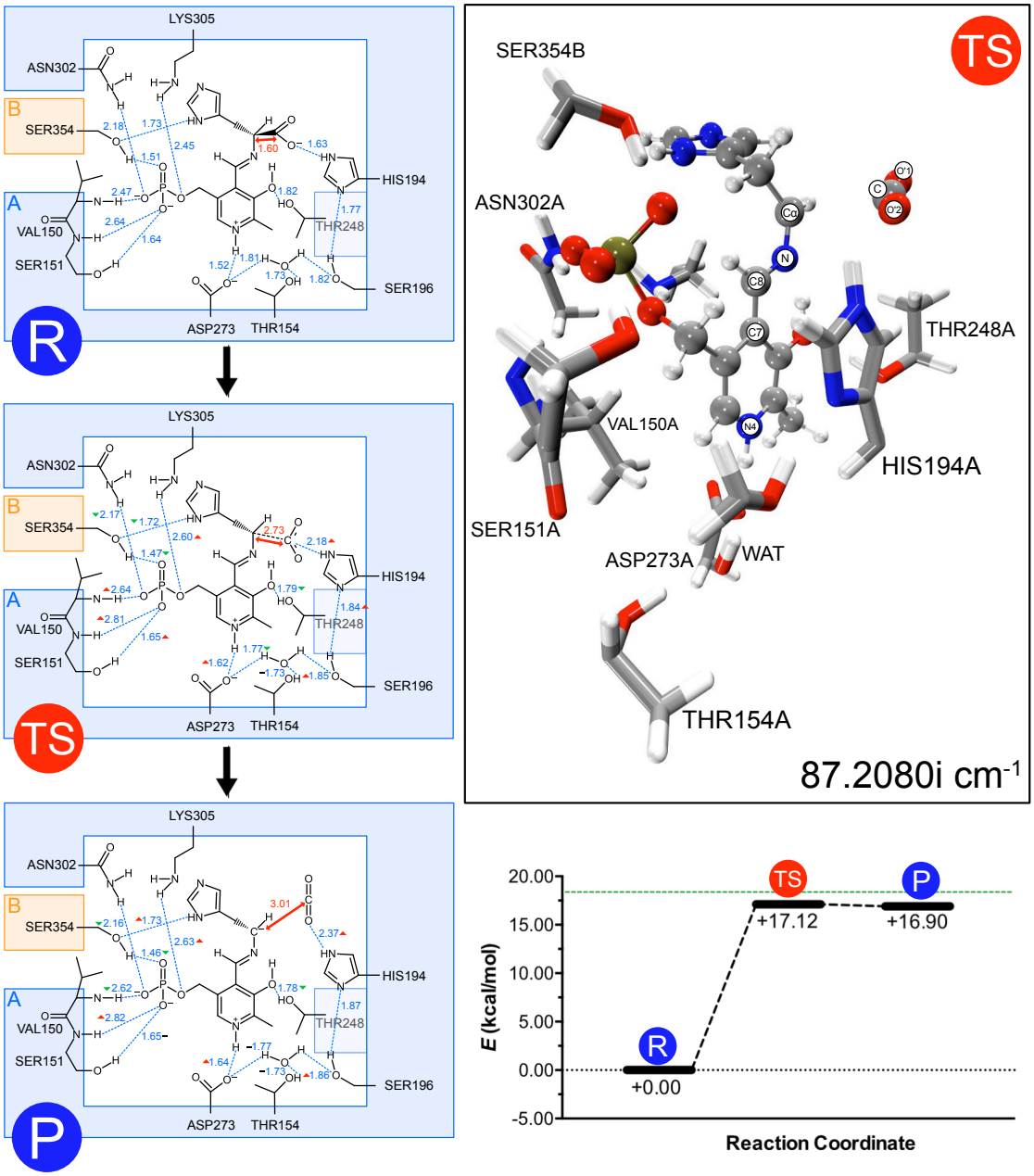



**Fig. 67 - Reactant (R), transition state (TS) and product (P) structures and energy profile.**

(Left) Wedge-Dash representation of the **EA-S**, water molecule and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was increased in the scan, and respectively interatomic distance in Å. (Top Right) CPK and licorice representations of the **EA-S**, water molecule and HL amino acid residues at the TS. (Bottom Right) Energy profile comparing reactant, TS and product. The green dashed line represents the $E_a$ predicted by the $k_{cat}$.

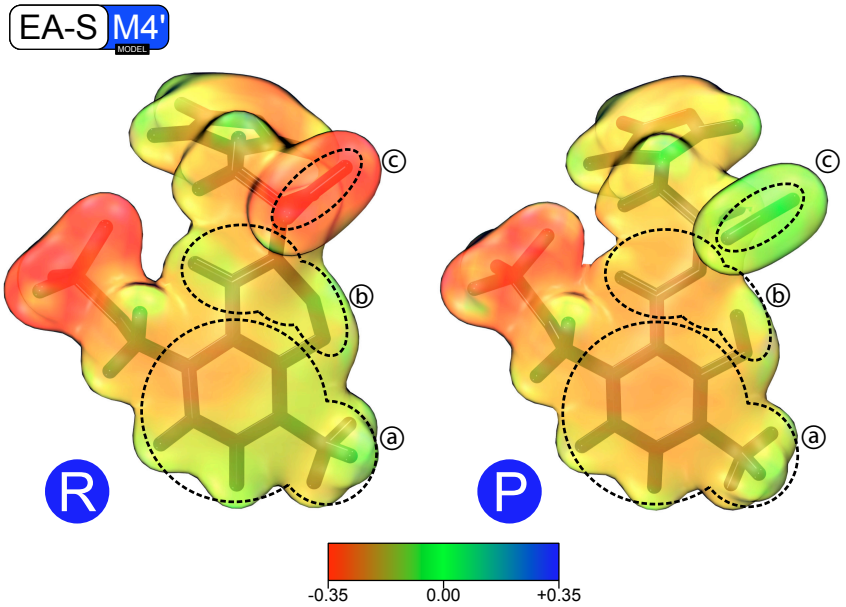



**Fig. 68 – Electrostatic map of reactant and product.**
Electrostatic maps were built based on Merz-Singh-Kollman charges. Red regions correspond to more negative regions whereas blue ones correspond to more positive regions. Green regions are related to neutral regions. The electrostatic map was built considering 1.76 points/Bohr$^3$.

**Table 5 – Merz-Singh-Kollman charges (atomic units) assigned to EA-S, respective regions a and b, carboxylate group/carbon dioxide (region c), phosphate group of PLP and Asp-273A for each reactional state (reactant, transition state and product).**

| Structure | Charge (a.u.) | | |
|---|---|---|---|
| | Reactant (R) | Transition State (TS) | Product (P) |
| EA-S | -1.180714 | -1.666394 | -1.722829 |
| Carboxylate Group/Carbon Dioxide | -0.623646 | -0.123422 | -0.055462 |
| PLP – Region a | +0.732749 | +0.354325 | +0.339217 |
| PLP – Region b | -0.568023 | -0.510279 | -0.507979 |
| Phosphate Group | -1.653409 | -1.686586 | -1.678557 |
| Asp-273A | -0.852647 | -0.868528 | -0.868990 |

This model (**EA-S-M4'**) provided for the first time an energy profile that suites very well to what is obtained with the experimental kinetic studies. Comparing with the previous models (**Fig. 69**), we have now a model that presents an activation barrier for the first step that is slightly lower than the one that is predicted by experimental studies.

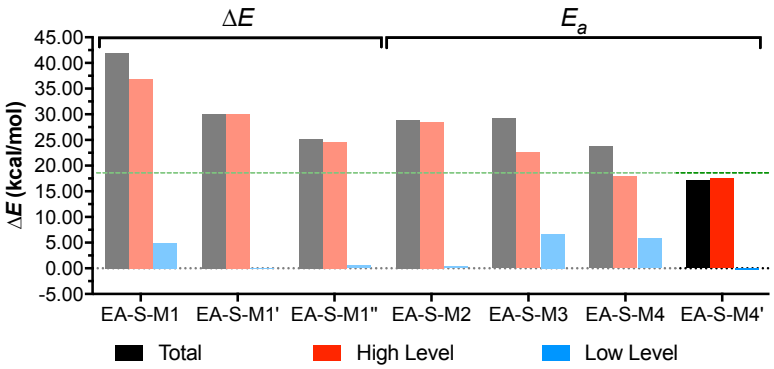



**Fig. 69 - Energy profile comparison between models.**
Black, red and blue bars correspond to the energy of all the system, HL layer and LL layer, respectively. The green dashed line represents the $E_a$ predicted by the experimental $k_{cat}$.

Until now, all the presented calculations only contemplate the electronic Hamiltonian, therefore a subsequent set of calculations were done aiming the nuclear Hamiltonian determination.

Frequency calculations on the **EA-S-M4'** model, using Gaussian09, were done using the obtained and optimized structures of reactant, TS and product. Results indicate a great increase of entropy when the TS is achieved, being even greater when the product is formed. This increase of entropy considerably decreases the energetic barrier by 3.63 kcal/mol and stabilizes the product by 5.01 kcal/mol. These results are in agreement with what was expected, since during the first step one molecule (external aldimine) is broken to generate two molecules (quinonoid intermediate and carbon dioxide). When this thermal correction is considered, we can talk about $\Delta G^{\ddagger}$ and $\Delta G_r$, the free Gibbs energy of activation and reaction respectively. Then, the $\Delta G^{\ddagger}$ for the first step is 13.48 kcal/mol and the $\Delta G_r$ is 10.88 kcal/mol, unraveling an exergonic step (**Fig. 70**).

As described before, DFT based methods have some difficulties in describing accurately intermolecular interactions, especially van der Waals forces. In order to overcome this situation, a complementary calculation using a dispersion correction functional (GD3) were done to get a better accounting of those intermolecular interactions. Considering the GD3 corrections, $\Delta G^{\ddagger}$ decreased to 13.38 kcal/mol and $\Delta G_r$ also decreased to 11.78 kcal/mol after GD3 corrections (**Fig. 70**). In this particular case, GD3 corrections only affect $\Delta G^{\ddagger}$ and $\Delta G_r$ by less than 1%, indicating a very acceptable description of non-bond interactions by the B3LYP functional.

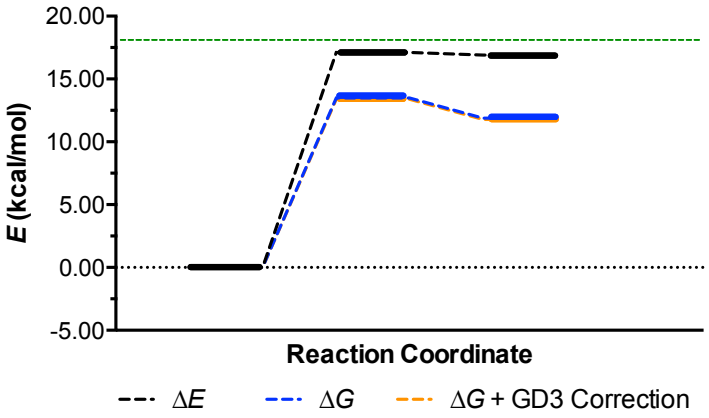



**Fig. 70 – Energetic profile of the first step of catalytic mechanism.**

In order to improve the theoretical description of the system, a benchmark of basis sets was performed to compare its influence on the final energies. To this end, different basis set from 6-31G(d) to a more complete 6-311++G(3df,2dp) were tested. Furthermore, a benchmark of functionals was also made in order to get the influence of the functionals used in the DFT calculations of the electronic energy. To this end, single-point energy (SP) calculations with hybrid functional that include a percentage of HF exchange were also done. In this case, M06 and M06-2X were used, and they have 27% or 54% of HF exchange term, respectively.

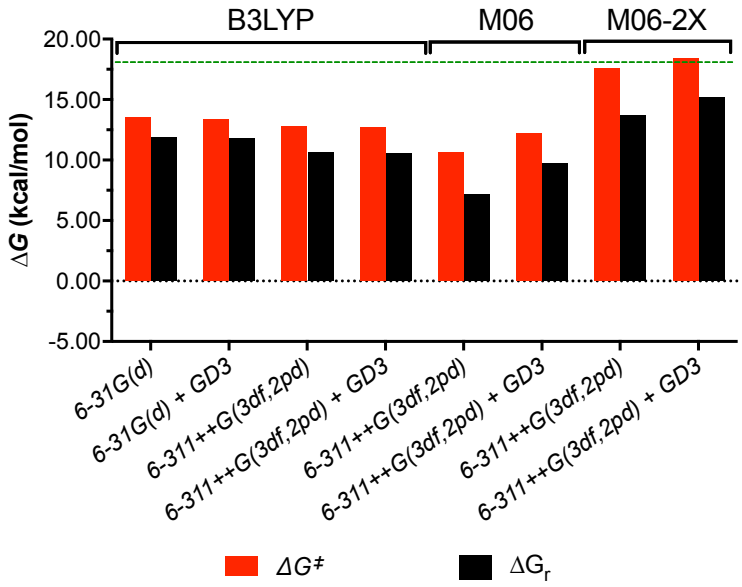



**Fig. 71 – Benchmark of functionals and basis sets for the first step, using EA-S-M4' model.**

The results from the benchmark calculations (**Fig. 71**) showed that the effect of the basis set do not provide an important variation on the calculated energies for the first step using the **EA-S-M4'** model. The energies calculated by the different functionals have however provided different results. Interestingly, no major differences on the calculated energies computed with the B3LYP and the M06 functionals are observed. In fact, the M06 functional provided better results, lowering down either the activation energy and the reaction energy. This can in part be explained taking into account the strong delocalization effect of the electrons that is observed in this reaction Therefore, the inclusion of the additional HF exchange by the M06 functional provided a better description of the system. Interestingly a major increase on the activation and reaction energies are observed when using the M06-2X functionals. This functional is characterized by adding more 27% of HF exchange in relation to M06. We believe therefore that the introduction of this "extra" effect is having a negative impact on calculated energies. The results would become similar to the ones provided by B3LYP and M06 functionals if the reactant, TS and product of the reaction were re-optimized with this functional. However, this would lead to extremely high computational cost that was not feasible due to disk and memory problems on the available hardware.

### 3.2. Step 2 – Intermediate protonation

As mentioned above, the available experimental data suggest that the catalytic mechanism of mHDC is a two-step type of mechanism where the second step involves the protonation of the product from the first step, i.e., the quinonoid intermediate (**QI** model) by an unknown proton-donor. Based on the site-directed mutagenic experiments two amino acid residues from the active site can play such role and they are Tyr-334B or/and His-194A. [66]

When Tyr-334B is mutated by glycine the enzyme becomes inactive. [66] This fact indicates that this amino acid residue is important for the catalytic process and can play a crucial role either in the first step or in the second step of the mechanism. Since the theoretical results from the previous section indicate that Tyr-334B is not required for the first step, it is therefore evident that it must be somehow involved in the second step of the mechanism. The MD simulations made in this work, reinforce this idea since this residue can be found very near to carbon Cα of the quinonoid intermediate, and therefore can be a good candidate to protonate this carbon.

Site directed mutagenesis of His-194A indicate that when this residue is mutated by glycine the enzyme is still active but the catalytic process is very slow. [66] This result shows that this residue may not be directly involved in catalysis as Tyr-334B but, might be important to assist the proton transfer or be important for the correct orientation or stabilization of the reaction intermediates inside the active site. The theoretical results obtained in the last section reveal that in the first step of the catalytic mechanism, this residue establishes an important hydrogen bond with the carboxylate group of the external aldimine that was shown to be important for the correct orientation of the carboxylate group during the decarboxylation process. However, this role does not fully justify the results obtained with the site-directed mutagenesis of this residue and therefore another role maybe played by His-194A in the second step of the catalytic process. The MD simulations show that this residue can also be found very near to carbon Cα of the quinonoid intermediate, and behave as a proton donor or behave as a proton switcher that facilitates the proton transfer from Tyr-334B, that is locate very close to His-194A. Taking into account the pKa of histidine, it can be disclosed the participation of this residue as a direct proton donor, and therefore its role as a proton switcher gains more credibility.

Based on the results obtained by the MD simulations, and taking into account these two mechanistic alternatives, two structures were selected and subjected to QM/MM protocol that have similar time prevalence along the MD simulation. The first model, **QI-M1A**, was used to test if the protonation of the quinonoid intermediate could

be made by Tyr-334B and mediated by His-194A. The second one, **QI-M1B** was used to test if the proton transfer occurs directly through Tyr-334B, without the participation of His-194A. Unfortunately, these models were built based on the MD simulations that were modeled accordingly to the initial structure available in the PDB file. Consequently, both models have the N-H bond pointing to the phosphate group of PLP instead to O1 oxygen as it was shown to be important in the first step described on the previous sections. Therefore, the calculations that were performed with the two mechanistic alternatives (**QI-M1A** and **QI-M1B**) will be briefly described without an extensive analysis. It should be noted here that these calculations were done at the same time that the ones from the first step were conducted. Therefore, at that time it was not known if the effect of rotating the NH group of the imine linkage would interfere with the catalysis or not. Nevertheless, the description of these calculations will be important to calculate the effect of the imine linkage rotation on the energies of the second step.

In order to simulate the second step with the imine linkage rotated, two other models were created. These models were built based on the product of the reaction from the **EA-S-M4'** model. These models were called **QI-M4A** and **QI-M4B**. They both have the same structural disposition of the **EA-S-M4'** model, but the protonation by Tyr-334B of the quinonoid intermediate was tested with the mediation of His-194A or without it, respectively.

### 3.2.1. Model QI-M1A: His-194A protonates the QI

This first model (**QI-M1A**) has the NH group of His-194A pointing towards carbon Cα of quinonoid intermediate. The hydroxyl group of Tyr-334B is also aligned with the other nitrogen atom of the imidazole ring of His-194A. The model has 93 atoms in the HL layer, and includes the quinonoid intermediate and the following amino acid residues: His-194A, Tyr-334B, Thr-248A, Asp-273A, Ser-151A, Ser-196A and Ser-354B (**Fig. 72**).

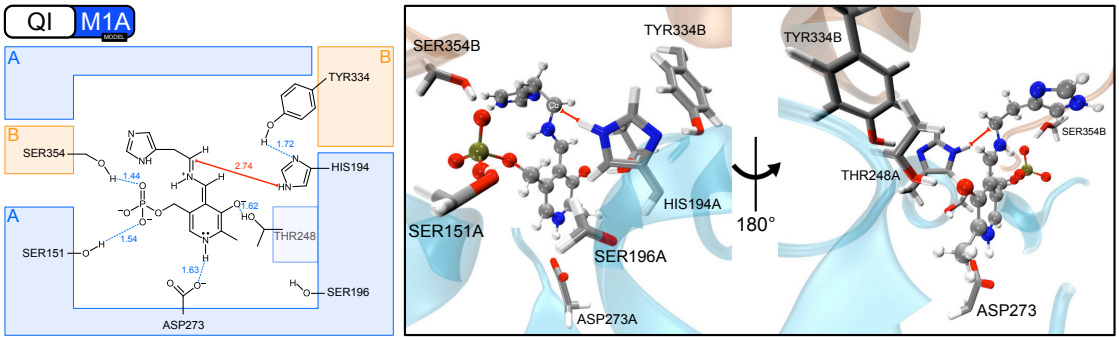



**Fig. 72 - Structure of the model QI-M1A.**
(Left) Wedge-Dash representation of the quinonoid intermediate (reactant) and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was decreased in the scan, and respectively interatomic distance in Å. (Right) CPK and licorice representations of the quinonoid intermediate and HL amino acid residues, respectively.

The theoretical results that were obtained with the **QI-M1A** model revealed that His-194A can behave as a proton switcher in the protonation of the carbon Cα of quinonoid intermediate by Tyr-334B. The reaction has a $E_a$ of 7.74 kcal/mol and the reaction is very exothermic ($E_r$ of -26.08 kcal/mol). The major contributions for the activation and reaction energies came from HL atoms, once the contribution of LL layer is +2.61 kcal/mol for the activation barrier and -1.18 kcal/mol for the energy of reaction. (**Fig. 73**)

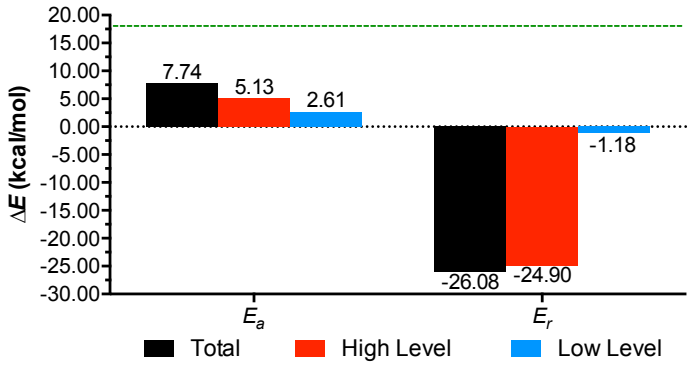



**Fig. 73 - Activation ($E_a$) and reaction ($E_r$) energies.**
Energies values considering all atoms of the system system (black), the HL atoms (red) or the LL atoms (blue). The green dashed line represents the $E_a$ predicted by the $k_{cat}$.

The TS of this reaction was confirmed by a unique imaginary frequency at 1431.24 $cm^{-1}$ (**Fig. 74**). During the protonation step, the proton from His-194A migrates to carbon Cα of the quinonoid intermediate and the opposite nitrogen of imidazole group of His-194A receives a proton from the hydroxyl group of Tyr-334B. At the end of the reaction, His-194A recovered its initial protonation state and the negative charge migrates from the quinonoid intermediate to Tyr-334B (**Fig. 74**). As seen before, Tyr-334B belongs to a very flexible and solvent-exposed loop which allows an easy stabilization of the generated negative charge and also promotes a quickly re-protonation by the solvent.

During this step, the N4-H bond length increased from 1.07 Å (reactant) to 1.11 Å (product) (**Fig. 74**), indicating the recovery of the positive charge of N4 nitrogen. This was a consequence of the protonation of carbon Cα of PLP by His-194A. Consequently, N4-H bond became weaker and this proton got a higher acidic character. Therefore, the interaction between this proton and the oxygen of Asp-273A became stronger, a phenomenon that is corroborated by a decrease of the interaction distance (reactant: 1.63 Å vs product: 1.51 Å) (**Fig. 74**).

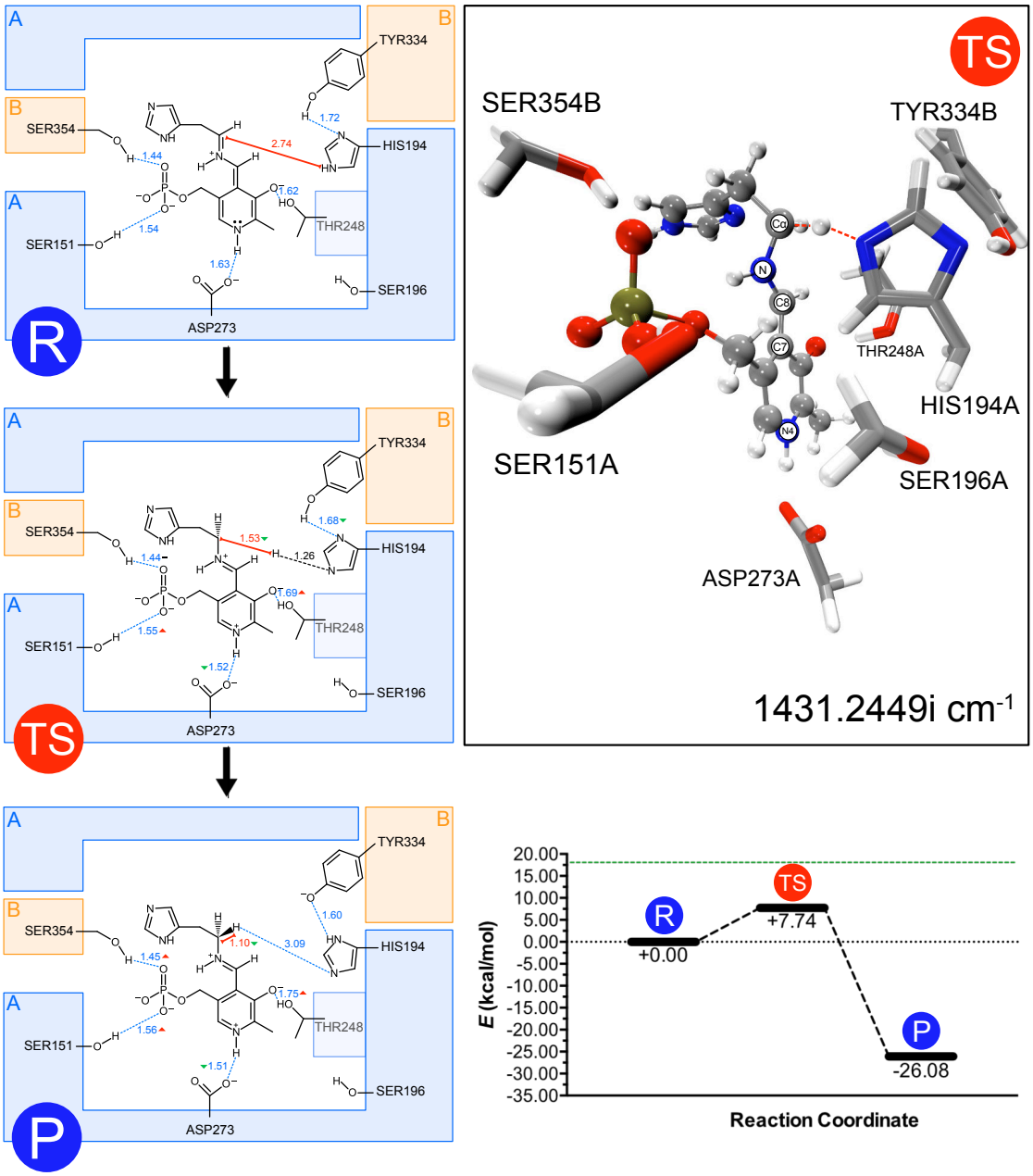



**Fig. 74 - Reactant (R), transition state (TS) and product (P) structures and energy profile.**
(Left) Wedge-Dash representation of the quinonoid intermediate and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was decreased in the scan, and respectively interatomic distance in Å. (Top Right) CPK and licorice representations of the quinonoid intermediate and HL amino acid residues at the TS. (Bottom Right) Energy profile comparing reactant, TS and product. The green dashed line represents the $E_a$ predicted by the $k_{cat}$.

Although this model (**QI-M1A**) is inaccurate, it showed that His-194A can protonate the quinonoid intermediate but it requires Tyr-334B to re-protonate His-194A simultaneously. We disclose here a key role of Tyr-334B in the protonation of quinonoid intermediate mediated by the His-194A. The next model (**QI-M1B**) describes a direct protonation of quinonoid intermediate exclusively promoted by the Tyr-334B.

### 3.2.2. Model QI-M1B: Tyr-334B protonates the QI

The next model (**QI-M1B**) describes a direct protonation of quinonoid intermediate exclusively promoted by the Tyr-334B. As mentioned before this model was built using a structure from the MD simulation and it is faithful to the original structure available in PDB file. The **QI-M1B** model has exactly the same amino acid residues in the HL layer comparatively as the previous model (**QI-M1A**). The difference lies in the reaction coordinate that was tested. In this case, the distance between carbon Cα and the proton of Tyr-334B was incrementally decreased, simulating the protonation of quinonoid intermediate by Tyr-334B. The structure used as reactant (**Fig. 75**) was obtained through an optimization using the QM/MM scheme B3LYP/6-31G(d):AMBER(ff99SB+GAFF).

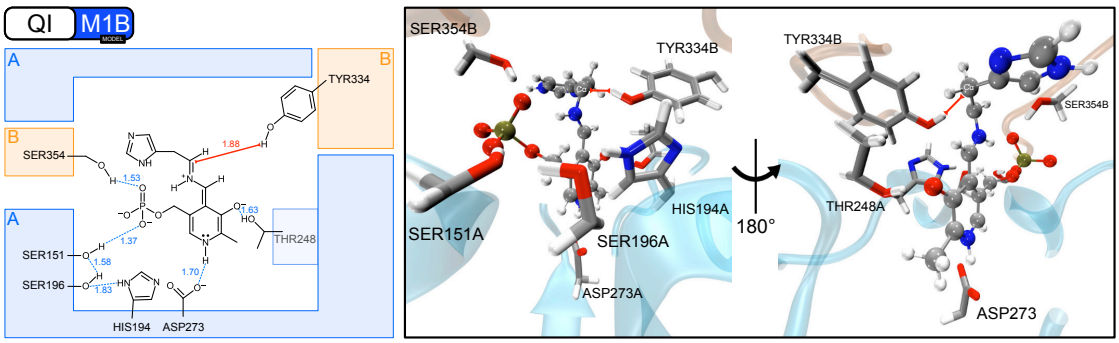



**Fig. 75 - Structure of the QI-M1B model.**
(Left) Wedge-Dash representation of the quinonoid intermediate (reactant) and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was decreased in the scan, and respectively interatomic distance in Å. (Right) CPK and licorice representations of the quinonoid intermediate and HL amino acid residues, respectively.

The theoretical results obtained with this model confirmed that the protonation of the quinonoid intermediate by Tyr-334B is possible and that the process is kinetically more favorable than the previous model (**QI-M1A**) where His-194A mediates the proton transfer. The activation barrier for this step is 0.38 kcal/mol (**Fig. 76**) and it is significantly lower than the barrier imposed to the protonation mediated by His-194A (7.74 kcal/mol (**Fig. 73**)). Nonetheless, this step is thermodynamically less favorable since the $E_r$ is -19.77 kcal/mol (**Fig. 76**), which makes the reaction less exothermic by 6.31 kcal/mol comparatively to the protonation catalyzed by His-194A (-26.08 kcal/mol (**Fig. 73**)).

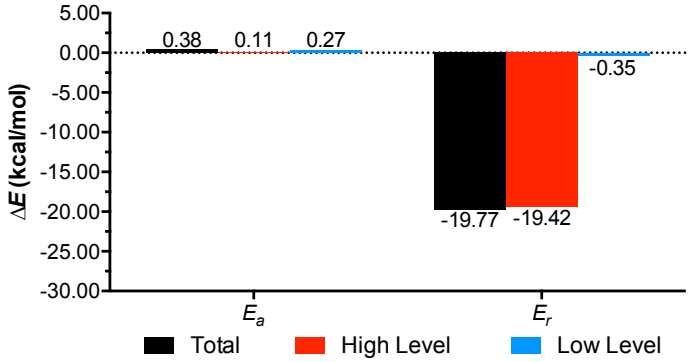



**Fig. 76 - Activation ($E_a$) and reaction ($E_r$) energies.**
Energies values considering all atoms of the system system (black), the HL atoms (red) or the LL atoms (blue). The green dashed line represents the $E_a$ predicted by the $k_{cat}$.

The TS of this step was confirmed by an imaginary frequency at 563.7731i $cm^{-1}$ (**Fig. 77**). During the protonation step, the Cα of the quinonoid intermediate was protonated by Tyr-334B with a very low activation barrier. During the reaction, the negative charge, that was initially located in the PLP cofactor, is delocalized to the Cα in the TS a fact that help the proton transfer from Tyr-334B. We can confirm this delocalization, since the N4-H bond length increased (reactant: 1.06 Å vs product: 1.08 Å) (**Fig. 77**) indicating a more positive N4 nitrogen. Accordingly, Asp-273A established a stronger non-bond interaction with the proton bonded to N4, once the interaction distance decreases from 1.70 Å to 1.64 Å (**Fig. 77**). With respect to the phosphate group, no significant changes were observed, corroborating the phosphate role as an anchor of PLP.

These results confirm the key role played by Tyr-334B in the protonation step of the quinonoid intermediate through a reaction that is very exothermic and also very fast (low activation energy). In the next models, this hypothesis was tested with models that accurately describe the quinonoid intermediate**,** once they correspond to the product of the first step, and they have the correct orientation of N-H bond.

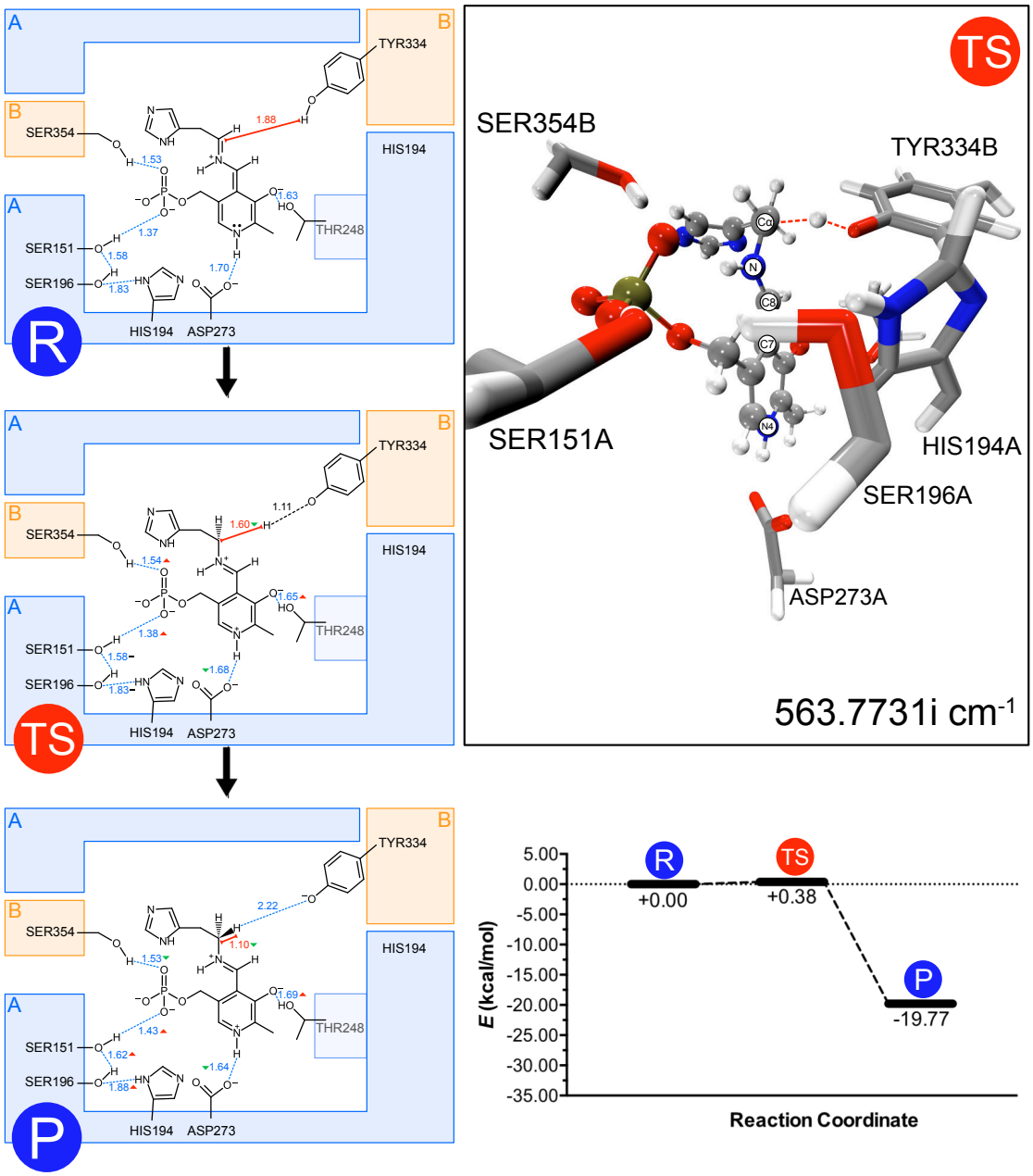



**Fig. 77 - Reactant (R), transition state (TS) and product (P) structures and energy profile.**

(Left) Wedge-Dash representation of the quinonoid intermediate and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was decreased in the scan, and respectively interatomic distance in Å. (Top Right) CPK and licorice representations of the quinonoid intermediate and HL amino acid residues at the TS. (Bottom Right) Energy profile comparing reactant, TS and product. The green dashed line represents the $E_a$ predicted by the $k_{cat}$.

### 3.2.3. Model QI-M4A: His-194A protonates the QI (N-H bond rotated)

As mentioned above, this model (**QI-M4A**) was built using, as initial structure, the product from the first step (**EA-S-M4'** model). The HL layer had the same amino acid residues of **QI-M1A** model plus Tyr-334B that has an important role for the quinonoid intermediate protonation. The complex was optimized and the resulting structure (**Fig. 78**) was used to run the linear scan of the reaction coordinate that involves the proton migration from His-194A to the carbon Cα of the quinonoid intermediate.

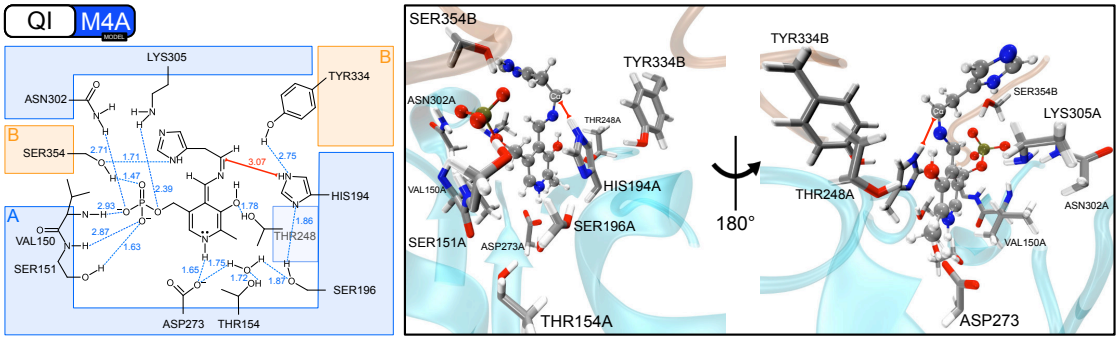



**Fig. 78 - Structure of the QI-M4A model.**
(Left) Wedge-Dash representation of the quinonoid intermediate and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was decreased in the scan, and respectively interatomic distance in Å. (Right) CPK and licorice representations of the quinonoid intermediate and HL amino acid residues, respectively.

The TS of this step was achieved using this model (**QI-M4A**) and confirmed by a unique imaginary frequency at 1517.7971 $cm^{-1}$ (**Fig. 80**). The activation barrier for the protonation of quinonoid intermediate by His-194A is 13.84 kcal/mol (**Fig. 79**). This barrier is similar to the one that was obtained using the **QI-M1A** model, but the N-H bond rotation induces an energy penalization in the **QI-M4A** model. The increase of the energy barrier is a consequence of the difficulties in protonation the Cα of the quinonoid intermediate, once His-194A is worst orientated to promote that reaction. Similarly to what was found in the QI-M4A model, the reaction is very exothermic with the **QI-M4A** model ($E_r$ = -16.80 kcal/mol) (**Fig. 79**).

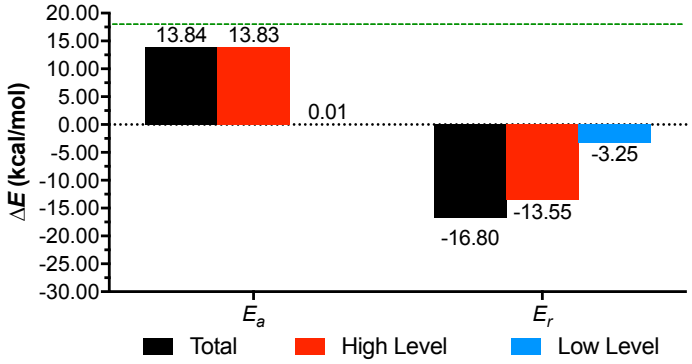



**Fig. 79 - Activation ($E_a$) and reaction ($E_r$) energies.**
Energies values considering all atoms of the system system (black), the HL atoms (red) or the LL atoms (blue). The green dashed line represents the $E_a$ predicted by the $k_{cat}$.

In this model, His-194A donates its proton to the quinonoid intermediate, and it becomes negatively charged (**Fig. 80**). The negative charge of His-194A is promptly stabilized by the Tyr-334B, which is corroborated by a decrease in the hydrogen interaction between Tyr-334B and His-194A (reactant: 2.75 Å vs product: 2.44 Å) (**Fig. 80**). When the protonation of the quinonoid intermediate occurs, PLP ring loses the negative charge that was accommodated during the first step. This evidence is confirmed by a strong interaction between nitrogen N4 and Asp-273A (reactant: 1.65 Å vs product: 1.53 Å), which results in a more positively charges nitrogen N4 (**Fig. 80**). Consequently, the water molecule nearby Asp-273A becomes more apart (reactant: 1.75 Å vs product: 1.80 Å) and the interaction of that water with other amino acid residues, Thr-154A and Ser-196A, becomes stronger (**Fig. 80**).

As all previous models have already showed that the phosphate group of PLP does not undergo significant changes in its interactions profile with the nearby amino acid residues (**Fig. 80**).

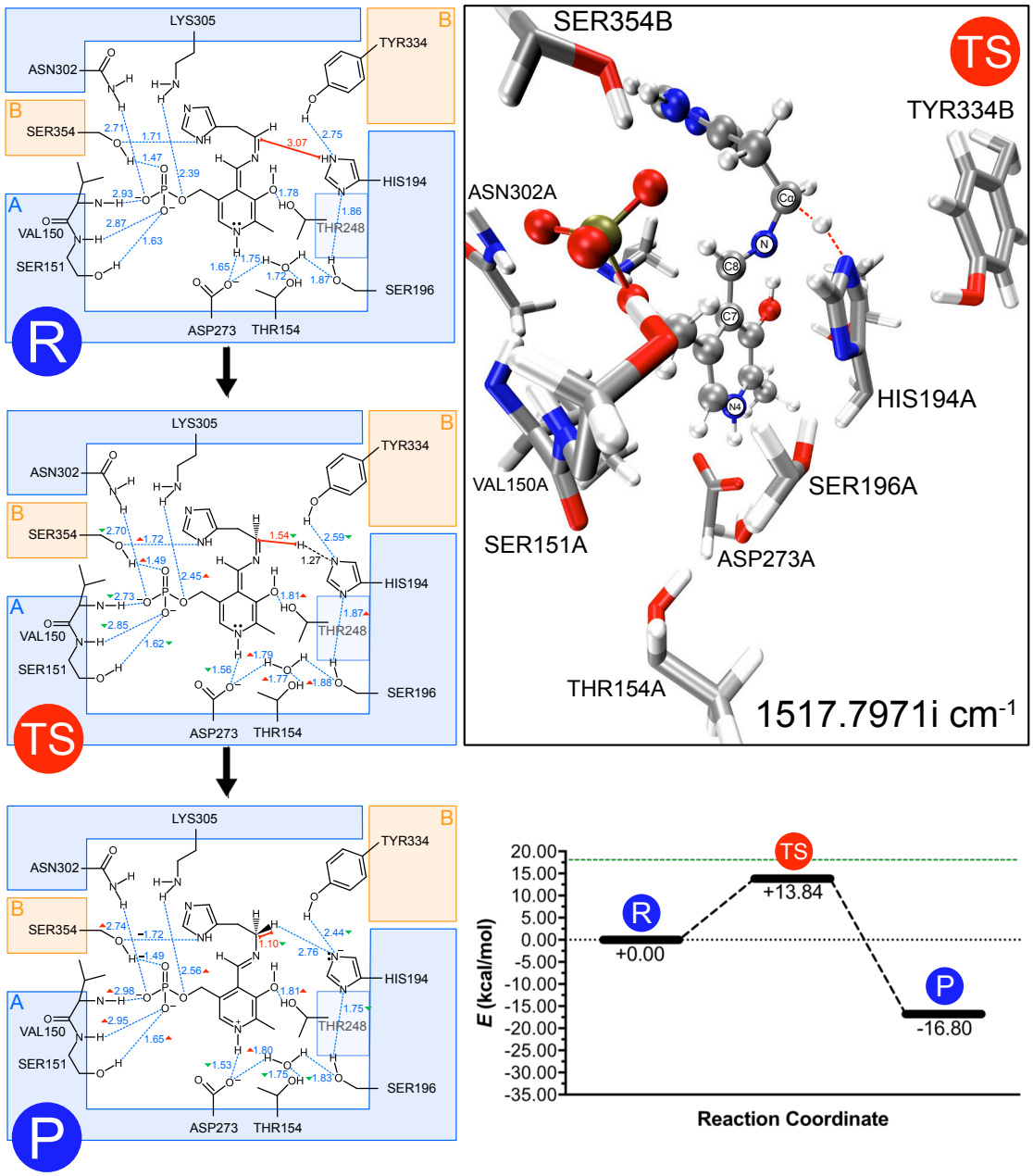



**Fig. 80 - Reactant (R), transition state (TS) and product (P) structures and energy profile.**
(Left) Wedge-Dash representation of the quinonoid intermediate and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was decreased in the scan, and respectively interatomic distance in Å. (Top Right) CPK and licorice representations of the quinonoid intermediate and HL amino acid residues at the TS. (Bottom Right) Energy profile comparing reactant, TS and product. The green dashed line represents the $E_a$ predicted by the $k_{cat}$.

The **QI-M4A** model describes a possible mechanism for the protonation of quinonoid intermediate. However, the activation of this step is very high taking into account that the first step of the mechanism that requires 13.48 kcal/mol for the first step. This would turn the full mechanism not feasible under physiological conditions and therefore the direct protonation of the quinonoid intermediate by Tyr-334B is more likely to occur. Mutagenesis studies confirm these results, once the His-194A mutants kept

their catalytic activity, besides the lower rate. It means that His-194A is important for the catalysis, but it is not involved in a major role during the reaction.

### 3.2.4. Model QI-M4B: Tyr-334B protonates the QI (N-H bond rotated)

Similar to the previous model (**QI-M4A**), the model (**QI-M4B**) was built having as reference, the product from the first step retrieved from the model **EA-S-M4'** model. This means that this model is equal to the previous one, but, in this case, we studied the direct protonation of carbon Cα of the quinonoid intermediate by Tyr-334B. The complex was optimized and the resulting structure (**Fig. 81**) was used to run the linear scan to follow the reaction coordinate that is involved in the proton migration from the hydroxyl group of Tyr-334B to carbon Cα of the quinonoid intermediate.

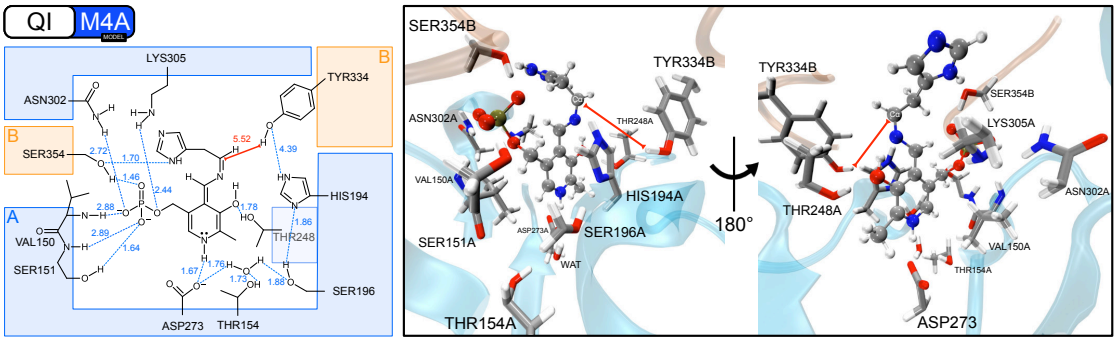



**Fig. 81 - Structure of the QI-M4B model.**
(Left) Wedge-Dash representation of the quinonoid intermediate (reactant) and the HL amino acid residues. Blue dashed lines represent relevant non-bonded interactions and respectively distance in Å. Red double arrow line represents the bond whose length was decreased in the scan, and respectively interatomic distance in Å. (Right) CPK and licorice representations of the quinonoid intermediate and HL amino acid residues, respectively.

Unfortunately, the full characterization of the TS and respective adjacent minima (reactant and product) for this reaction are still not completely characterized. Currently, the full convergence of the TS structure has not been fulfilled, but frequency calculation of the structure closest to the TS shows only one imaginary frequency at 25.8653i $cm^{-1}$ and coincides with the reaction coordinate that is being followed. This means that is just a matter of time to gather the full characterization of the TS of this step. Several structures of the reactant and the product of this reaction have already been obtained, but in order to obtain the ones that are truly connected to the TS structure, we need to wait for the full characterization of the TS and obtain them through IRC calculations. At this point only global energies regarding the activation and reaction energies can be presented with the B3LYP/6-31G(d) level of theory. However, we do believe that the final

energies using the same level of theory will not change significantly to the ones that are presented here.

Analogously to what happenned to the **QI-M1B** model, we confirmed a very fast reaction for the protonation of the quinonoid intermediate, since the activation barrier is very low ($E_a$ = 2.87 kcal/mol). Moreover, this model also confirmed a high exothermic reaction ($E_r$ = -26.02 kcal/mol), something that has already been observed with the **QI-M1B** model. These results demonstrate that Tyr-335B is the proton donor that protonates carbon Cα of the **QI** intermediate. This result goes in line with the mutagenic experiments that reveal that if this residue is mutated by a glycine the enzyme becomes inactive. His-194A has a secondary role in this process and is not directly involved in the protonation step. However, it plays a very important role to guide Tyr-334B nearby carbon Cα of the quinonoid intermediate, and, at the same time, to stabilize the negatively charged that is kept by Tyr-334B after the protonation step. Such stabilization also helps, in the product of the reaction, to move Tyr-334B towards the solvent, in order to become re-protonated by the water molecules, and by this way promote enzymatic turnover. The role played by His-194A in the reaction is, therefore, not so innocent has it was previously thought.

The obtained results so far also indicate some insight about the electronic delocalization along the PLP cofactor during this step. As we observed in the first step (**EA-S-M4'**), the decarboxylation process produces a negative charge at carbon Cα, that is deeply stabilized by the PLP ring, (region a). Since the product of the first step is the reactant of the second one, we confirmed the same phenomenon in the reactant of this step (**QI-M4B** model) (**Fig. 82**). When the proton, donated by Tyr-334B, moves towards carbon Cα, the electronic density is withdrawn from the PLP ring (region a) to the carbon Cα (region c) in order to promote the nucleophilicity of carbon Cα, and, by this way, to become protonated (**Fig. 82** and **Table 6**). We observed that in this reaction the PLP ring (region a) becomes more positively charged and the atoms surrounding nitrogen N (region b) also becomes less negative. This results into an increase of the global charge of the new external aldimine that is formed after the protonation step (reactant: -1.786819 vs product: -1.045458).

In sum, the PLP cofactor is very important to stabilize the transient negative charge that is formed after the decarboxylation step. Afterward, the quinonoid intermediate is generated, and the PLP cofactor recovers its initial electronic stage.

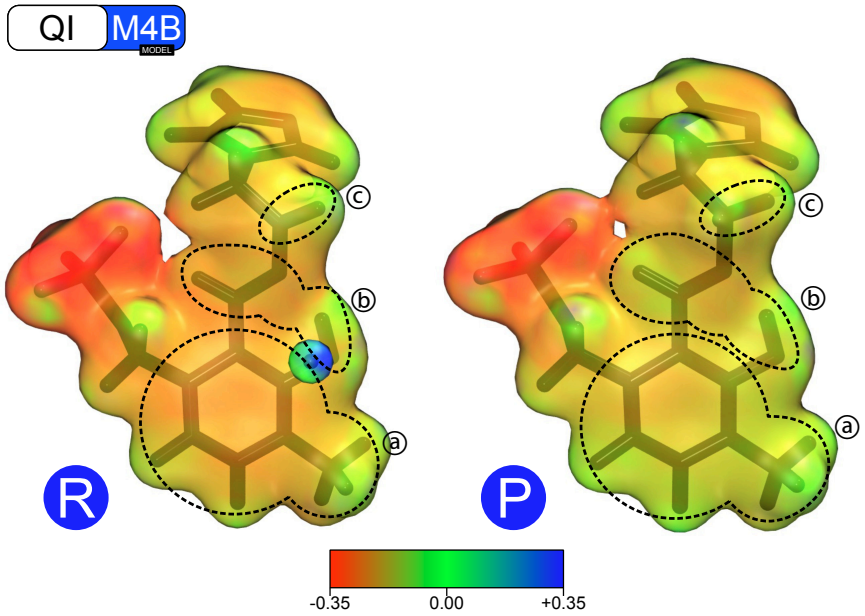



**Fig. 82 - Electrostatic map of reactant and product.**

Electrostatic maps were built based on Merz-Singh-Kollman charges. Red regions correspond to more negative regions whereas blue ones correspond to more positive regions. Green regions are related to neutral regions. The electrostatic map was built considering 1.76 points/Bohr$^3$.

**Table 6 - Merz-Singh-Kollman charges (atomic units) assigned to QI, respective regions a and b, carboxylate group/carbon dioxide (region c), phosphate group of PLP and Asp-273A for reactant and product).**

| Structure | Charge (a.u.) | |
|---|---|---|
| | Reactant (R) | Product (P) |
| QI | -1.786819 | -1.045458 |
| Region c | +0.333209 | +0.527512 |
| PLP – Region a | +0.341382 | +0.717479 |
| PLP – Region b | -0.5983585 | -0.300780 |
| Phosphate Group | -1.738189 | -1.702650 |
| Asp-273A | -0.857459 | -0.858841 |

# D. Conclusions and Future Perspectives

*Science never solves a problem*
*without creating ten more.*
**George Shaw**

This page was intentionally left blank.

HDC is currently an important drug target that can be used in the treatment of several diseases, such as atopic dermatitis, allergies, CHF and cancer. This enzyme requires the presence of the PLP cofactor that is necessary for the catalytic process. Currently, there is only one structure of mHDC available on the PDB, and it was co-crystallized with an inhibitor bonded to the PLP cofactor. This structure together with other experimental studies generated several proposals about the catalytic mechanism of this enzyme.

There is a general consensus that the mechanism involves two steps. The first step involves the decarboxylation of the natural substrate and the concomitant release of carbon dioxide. The second step involves the protonation of the quinonoid intermediate by an active site residue. The available site-directed mutagenic studies also reveal that several active site residues are essential for catalysis: Asp-273A, Lys-305A, Tyr-334B, His-194A, Ser-304A, Thr-248A, Asn-302A, Tyr-80A, Ser-354B and Ala-83A. [66] The mutation of any of these residues either turns the enzyme inactive or decreases considerable the catalytic process. In spite of these results, it is still no fully understood how the catalytic mechanism takes place. For instance, nothing is known about which amino acid residue from the active site is involved in the last protonation step. It is also unknown the role played by His-194A, Tyr-334B and Asp-273A on catalysis that have been shown to be detrimental for the catalytic process. In addition, the only PDB structure that is available for this enzyme, includes the PLP cofactor bonded to substrate analogue with an unusual conformation of it that has never been found in other PLP-dependent enzymes. Either this conformation is an artifact generated by the crystallization process or from the inhibitory process. Otherwise this structure can also suggest that this enzyme employs a different type of mechanism that has never been observed in other PLP-dependent enzymes.

All of these questions remain unanswered and required further explanations. This is important since HDC is an important drug target and there is an urgent need to develop new inhibitors targeting this enzyme. In addition, understanding the catalytic mechanism of this enzyme will allow to understand the specificity of this enzyme in relation to the other enzymes of the same family. This is particularly important since AADC (another enzyme of this family) share many similarities to this enzyme but accepts a different type of substrates. Taking into account that AADC is another important drug target, the understanding of the differences related to their substrate specificity will help to develop new and more specific inhibitors targeting each of these enzymes.

In this work, theoretical and computational means were used to shade some light to each of these questions and also to describe with atomic detail the full catalytic mechanism of mHDC.

Since there is only one structure of mHDC available on the PDB, our first task focused on modeling the wild-type enzyme with the natural substrate bonded to it. This was particular important since the inhibitor adopts a different conformation in relation to the PLP cofactor that is not common among the other PLP-dependent enzymes. The generated model was then subjected to a MD simulation in order to optimize the structure of the full complex and accommodate all the changes that were introduced manually. This initial task was particular important to achieve all the milestones that were highlighted before since it allowed to sample several conformations of the PLP cofactor together with the substrate bonded to it, as well as the conformation of several active site residues and surrounding water molecules that were later one identified as crucial for the catalytic process.

Several models were then taken from the MD simulation, and the QM/MM models were built and used to study the catalytic mechanism of mHDC, using the ONIOM subtractive scheme and the DFT-B3LYP/6-31G(d):AMBER(ff99SB + GAFF) level of theory. In this process, several reaction coordinates were followed for each step that were used to explore different mechanistic proposals. It is worth mentioning here that the first step of the catalytic mechanism was calculated with a higher theoretical level of theory and with a more complete basis set (DFT-M06/6-311+G(d,p)/GD3 :AMBER(ff99SB + GAFF). However, since the final calculations of the second step are still not complete, we opted to just present in the conclusion of this thesis the energies that were obtained with lower theoretical level of theory. However, the results from the first step indicate that the results obtained with the higher theoretical level and a more complete basis set does not influence dramatically the energies that are presented in this section.

The QM/MM calculations revealed that the decarboxylation was impossible under physiological conditions using the complex of the wild-type enzyme with the natural substrate bonded to the PLP cofactor in the unusual conformation, similar to what is found in the PDB structure with the code 4E1O [17] ($E_a$ > 20 kcal/mol). Several modifications to the complex were introduced to explore different conformations of the active site residues, as well as different QM/MM schemes but with no success. These results clearly indicated that the unusual disposition of the natural substrate in relation to the PLP cofactor is not feasible for catalysis.

The wild-type enzyme model was then modeled with the substrate bonded to the PLP cofactor in a similar conformation that is observed in the majority of the PLP-dependent enzymes. This new configuration allows the best orientation of the imine bond, and allows the formation of an important hydrogen bonds between nitrogen N4 and the carbonyl groups of the PLP cofactor. In addition, such configuration promotes the positioning of the imidazole group of the substrate in a manner that endorses a hydrogen bond formation with Ser-354B that noticeably affects the decarboxylation process. In addition, in such model, His-194A becomes parallel aligned to the PLP cofactor and establishes an important hydrogen bond with the carboxylate group that was found to be important for the correct orientation and accommodation of the external aldimine inside the active site. In addition, His-194A also becomes aligned with Tyr-334B that was found to be important for the second step of the mechanism.

The QM/MM calculations of this model system showed that the catalytic mechanism is complete in two steps. The computed energies reveal that the decarboxylation is the rate-limiting step. It also indicates that the energetic profile for the catalytic mechanism of mHDC is highly endothermic reaction in the decarboxylation process, and highly exothermic reaction in the second step that involves the protonation of the quinonoid intermediate (**Fig. 83** black line).

The final energies of the energetic profile present in **Fig. 83** were computed with the DFT-B3LYP/6-31G(d):AMBER(ff99SB + GAFF) level of theory, including the thermal correction calculated with the DFT-B3LYP/6-31G(d):AMBER(ff99SB + GAFF) level of theory and the dispersion effect of the DFT calculations with the GD3 basis set.

The thermal corrections provide a good overview about the entropic contribution for the total energy along the calculated energetic pathway. The entropy contribution lowers either the activation barrier by 2.21 kcal/mol and the reaction energy by 4.16 kcal/mol of the first step (the entropic effects on the second step are still not available). The entropic effect observed in the first step can be explained by the fact that during the decarboxylation process, two molecules (quinonoid intermediate and carbon dioxide) are formed from only one molecule, the external aldimine. Consequently, the stabilization of the product from the first step is significantly more stabilized when the entropic contributions are considered (**Fig. 83** dark blue line).

The light blue line at the **Fig. 83** corresponds to the activation and reaction energies ($\Delta E$) for the second step, since the full characterization of the TS and minima were not achieved yet in this step, as explained before.

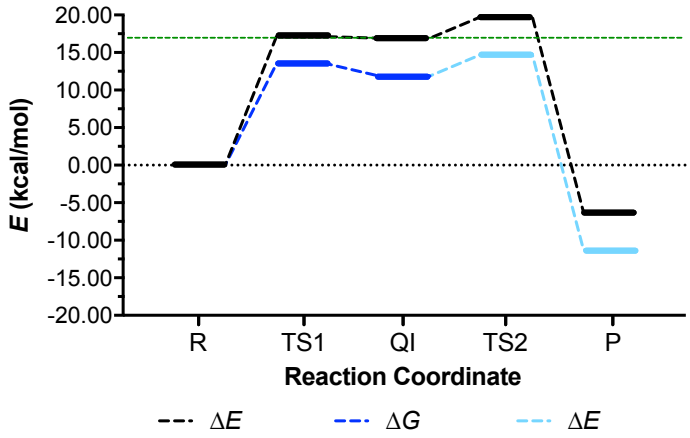



**Fig. 83 – Energetic profile of all reaction catalyzed by mHDC.**

The obtained results go in line with the experimental free energy that indicate that the rate-limiting step should be below 18 kcal/mol and in our results the rate-limiting step requires an activation energy of 13.48 kcal/mol.

The computed energies also provide interesting results regarding the different contribution of HL and LL layers in the QM/MM scheme that was used, in the activation and reaction energies of the decarboxylation step. The energetic profile for the first step (**Fig. 84**) reveals that the LL layer slightly stabilizes the TS and product as we expected. The major contribution for the energies are assigned to the HL layer that includes all the amino acids players involved in the decarboxylation process.

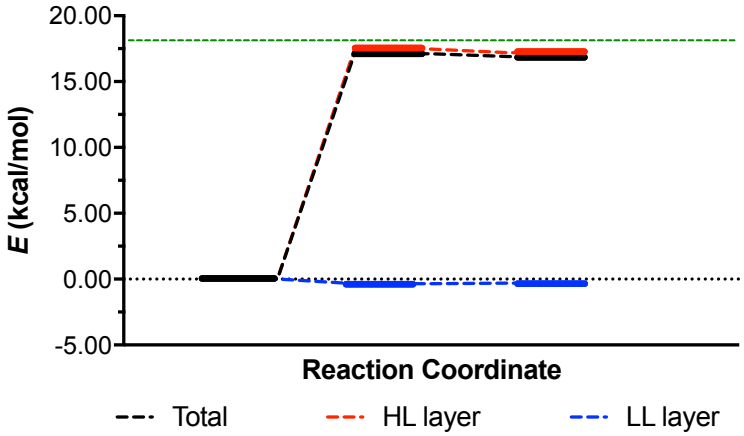



**Fig. 84 – Energetic profile of the decarboxylation process catalyzed by mHDC.**
Energies values considering all atoms of the system (black), the HL atoms (red) or the LL atoms (blue). The green dashed line represents the $E_a$ predicted by the $k_{cat}$.

The computed mechanism is also in good agreement with the available experimental data [66] that points Lys-305A, Asp-273A, Tyr-334B and His-194A as important residues for catalysis.

Lys-305A is crucial for the catalysis because it establishes an imine bond with the PLP cofactor, originating the internal aldimine. If the Lys-305A is not present, the internal aldimine is not formed, and, consequently, the decarboxylation of *L*-histidine is blocked [51, 52].

The computed catalytic mechanism has shown that when Tyr-334B is directly involved in the protonation step of the second step of the catalytic mechanism the reaction is about one hundred million times faster than when His-194A is used to mediate the proton transfer. This explains why the mutation of Tyr-334B by a glycine turns the enzyme inactive whereas the mutation of His-194A by glycine only decreases the catalytic activity of the enzyme, but does not prevent the reaction to occur.

The role played by His-194A in the mechanism has also been disclose. His-194A is important to guide the tyrosine residue during its approximation to the carbon Cα of the quinonoid intermediate, and to guide Tyr-334B back to the solvent in order to be re-protonated. In addition, this residue is important in the first step of the mechanism to align the carboxylate group and the PLP ring along the decarboxylation process.

The mutagenesis studies also point Asp-273A as important for catalysis. The results obtained in this work revealed that this residue stabilizes the positively charged nitrogen N4 of the PLP cofactor, which promotes the charge withdrawing along the PLP cofactor and that helps the catalytic process (**Fig. 85**). However, such stabilization is not so straight as initial it can be thought. Indeed, when we mutated this residue by an alanine and repeated the first step of the catalytic mechanism (results not shown in this thesis), we found that the reaction is much faster and the reaction is not so endothermic as it was obtained when Asp-273A is present in the active site. This indicates that Asp-273A is not favoring the decarboxylation process, and this is an important feature for the full catalytic process. Indeed, if the negative charge was fully stabilized along the PLP cofactor, it would no longer be available for the protonation step of the second step that is required for the enzymatic turnover. In sum, we have a commitment that allows an overall faster reaction that allows the decarboxylation and protonation processes, besides the penalization of the first one.

The role played by Asp-273A is also helped by the network of hydrogen bonds that it established with the neighbor water molecule, Ser-196A and His-194A. When the external aldimine is formed, His-194A has also an important role because it establishes a hydrogen bond with the carboxylate group of the external aldimine. This interaction is important to orientate the carboxylate group, and to stabilize its negative charge. Consequently, the strong interaction between His-194A and the carboxylate group of PLP drags all the hydrogen bond network that is established between Asp-273A and His-194A, that includes a serine residue (Ser-196A) and a water molecule. In the reactants of the first step, the water molecule nearby Asp-273A is moved away and interacts weakly with the aspartate residue. This scenario is inverted after the decarboxylation process because His-194A loses its interaction with the carboxylate group, and it interacts weakly with Tyr-334B, which is not negatively charged. The water molecule placed nearby Asp-273A starts to interact with it in a stronger manner, stabilizing the negative charge of Asp-273A. This phenomenon is a consequence of weakening of the hydrogen bond between Asp-273A and nitrogen N4, which becomes much more negative than in the initial structure.

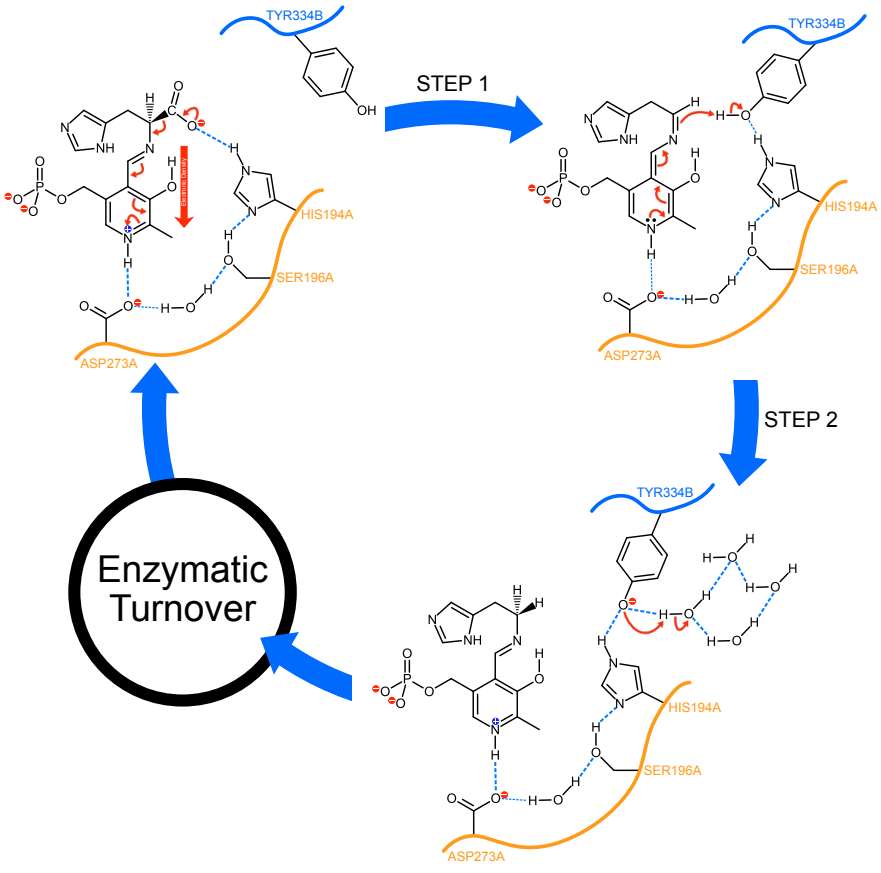



**Fig. 85 – Importance of the hydrogen bond network established between the carboxylate group and Asp-372A.**

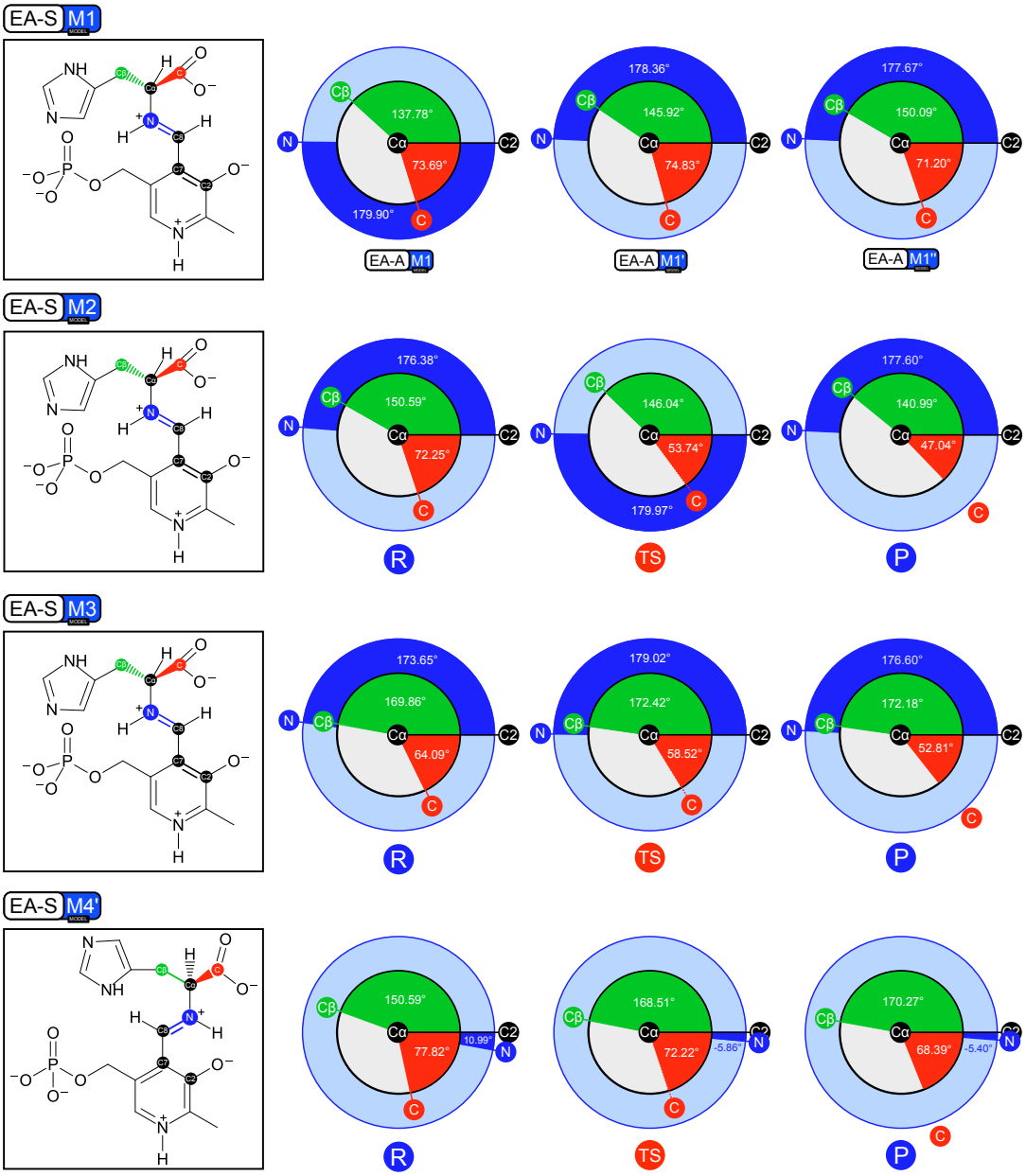



**Fig. 86 – Dihedral analysis of all tested model for the first step.**

As we saw, His-194A has several minor roles during catalysis, and one of them is to correct the orientation of the carboxylate group of the external aldimine during the decarboxylation process. This evidences the dihedral angle of the carboxylate group relatively to the PLP cofactor as an important feature for the catalysis. Dunathan [60, 61] have already proposed that during the decarboxylation process, the bond that is broken should be held at 90° to the plane in the relation the extended conjugated system. Accordingly, this positioning helps a faster catalysis for the decarboxylation step. [60, 61] We compare the dihedral angles of each model that was built to describe the decarboxylation step (**Fig. 86**) to disclose about the main differences. The model that

actually describes accurately the first step (**EA-S-M4'**) has the carboxylate group in almost perpendicular plane to the extended conjugated system of PLP (77.82°). If we look to the TS structures, we can confirm this trend, since **EA-S-M4'** model presents the carboxylate group closer to 90° comparatively to the other models. In this comparison, we cannot include the **EA-S-M1** models because they did not allow a TS characterization and the reaction was block by the proximity of Tyr-334B. In conclusion, Dunathan's theory was confirmed based on our models, but the carboxylate group does not need to be necessarily perpendicular to the conjugated system of PLP. Moreover, the correct orientation of the bond that is broken is not the only requirement to become the decarboxylation possible. Ser-354B is another amino acid residue that plays an important role in anchoring the phosphate group of PLP, but it is also crucial for the specificity of mHDC. If we overlap the active sites of AADC and mHDC, we can only disclose the absence of this serine residue in the AADC. In this work, we showed that Ser-354B establishes a strong hydrogen bond with the imidazole group of the substrate, and it gives the specificity of mHDC to *L*-histidine. For example, one of the substrates of AADC, *L*-DOPA, cannot bind the active site of mHDC due to the bigger side chain ring (dihydroxybenzene) that cannot be accommodate inside the active site because of stereochemical impediments. Then, we confirmed the role of Ser-354B in the substrate specificity of mHDC, and we can also predict a similar catalytic mechanism for AADC, since the players and their arrangement inside the active site is identical to those who have been identified in mHDC.

To finalize this conclusion, the complete description of the entire catalytic mechanism of mHDC will be presented (**Fig. 87**). The initial steps that are common to every PLP-dependent enzyme have already been described [51, 52], and the specific steps of the catalytic mechanism of mHDC were disclosed with this work.

The decarboxylation of *L*-histidine by the mHDC involves three major stages: i) the activation of mHDC by internal aldimine formation; ii) the transimination reaction, where the substrate binds the PLP cofactor; and iii) the decarboxylation reaction, where the substrate is decarboxylated and, then, protonated.

In the first stage, the PLP cofactor binds the ε-amino group from Lys-305A, originating an internal aldimine. The internal aldimine is the structure where the Lys-305A is linked to the PLP cofactor through an imine bond. When the enzyme has the internal aldimine formed in its binding pocket, it is ready to receive the substrate. Then, in the second stage, the internal aldimine binds the *L*-histidine (substrate) to form a gem-diamine intermediate. At this point, both Lys-305A and the substrate are bonded to the

PLP cofactor. Afterward, the Lys-305A releases the PLP cofactor, and this originates an external aldimine. The external aldimine is formed through the imine linkage between the substrate and the PLP cofactor. Since the external aldimine is formed, PLP ring withdraws the electronic density, and the bond that links the carboxylate group to the substrate becomes weaker. Consequently, carboxylate group is released in the form of carbon dioxide, and the resulting carbanion is stabilized by the PLP cofactor, originating a quinonoid intermediate. The quinonoid intermediate is, then, protonated by the hydroxyl group of Tyr-334B at the same position from where the carboxylate group was removed. Consequently, a new external aldimine is formed, and it consists in the product (histamine) bonded to the PLP cofactor. This new external aldimine enters again in the transimination stage, where the enzymatic turnover takes place. At this time, the histamine is released through the binding of Lys-305A and the internal aldimine is recovered for a new catalytic cycle.

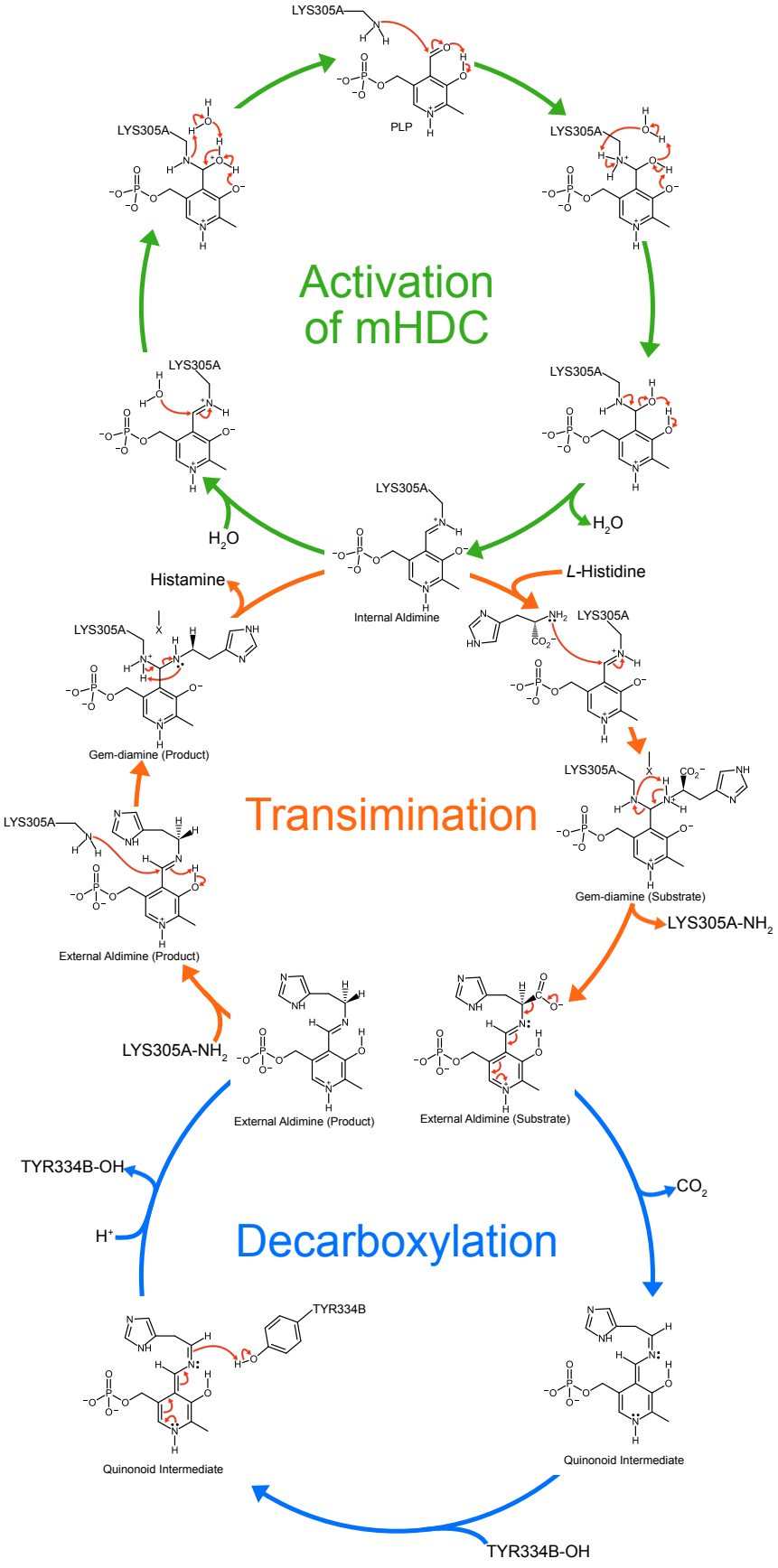



**Fig. 87 – Schematic representation of the proposal catalytic mechanism of mHDC.**

# F. Bibliography

This page was intentionally left blank.

1. Code, C.F., *Histamine and gastric secretion: a later look, 1955-1965.* Fed Proc, 1965. **24**(6): p. 1311-21.
2. Kahlson, G. and E. Rosengren, *New approaches to the physiology of histamine.* Physiol Rev, 1968. **48**(1): p. 155-96.
3. Shepherd, D.M. and D. Mackay, *The histidine decarboxylases.* Prog Med Chem, 1967. **5**: p. 199-250.
4. Hollis, T.M. and R.A. Ferrone, *Effects of shearing stress on aortic histamine synthesis.* Exp Mol Pathol, 1974. **20**(1): p. 1-10.
5. Alcaniz, L., et al., *Histamine production by human neutrophils.* FASEB J, 2013. **27**(7): p. 2902-10.
6. Pearce, F.L., *Biological effects of histamine: an overview.* Agents Actions, 1991. **33**(1-2): p. 4-7.
7. David E. Golan, A.H.T., Ehrin J. Armstrong, April W. Armstrong, *Principles of Pharmacology - The Pathophysiologic Basis of Drug Therapy*. Third Edition ed. 2001: LIPPINCOTT WILLIAMS & WILKINS.
8. Saxena, S.P., et al., *Histamine is an intracellular messenger mediating platelet aggregation.* Science, 1989. **243**(4898): p. 1596-9.
9. Glavin, G.B. and L.J. Brandes, *A novel non-H1, non-H2 histamine antagonist protects against cysteamine-induced duodenal ulcers in rats.* Pharmacology, 1988. **37**(5): p. 277-80.
10. Vassallo, P.F., et al., *The left ventricular contractility of the rat heart is modulated by changes in flow and a1-adrenoceptor stimulation.* Brazilian Journal of Medical and Biological Research, 1998. **31**: p. 1353-1359.
11. Smith, E.M., et al., *Corticotropin Releasing Factor (CRF) Activation of NF-κB-Directed Transcription in Leukocytes.* Cellular and Molecular Neurobiology, 2006. **26**(4): p. 1019-1034.
12. Kaltschmidt, B., et al., *NF-κB Regulates Spatial Memory Formation and Synaptic Plasticity through Protein Kinase A/CREB Signaling.* Molecular and Cellular Biology, 2006. **26**(8): p. 2936-2946.
13. Ai, W., et al., *Regulation of L-histidine decarboxylase and its role in carcinogenesis.* Prog Nucleic Acid Res Mol Biol, 2006. **81**: p. 231-70.
14. Haas, H.L., O.A. Sergeeva, and O. Selbach, *Histamine in the nervous system.* Physiol Rev, 2008. **88**(3): p. 1183-241.
15. Hakanson, R., *Mammalian histidine decarboxylase: interaction between apoenzyme and pyridoxal-5'-phosphate.* Eur J Pharmacol, 1967. **1**(5): p. 381-90.
16. Ellenbogen, L., E. Markley, and R.J. Taylor, Jr., *Inhibition of histidine decarboxylase by benzyl and aliphatic aminooxyamines.* Biochem Pharmacol, 1969. **18**(3): p. 683-5.
17. Komori, H., et al., *Structural study reveals that Ser-354 determines substrate specificity on human histidine decarboxylase.* J Biol Chem, 2012. **287**(34): p. 29175-83.
18. Mamune-Sato, R., et al., *Histidine decarboxylase in human basophilic leukemia (KU-812-F) cells. Characterization and induction by phorbol myristate acetate.* Biochem Pharmacol, 1990. **40**(5): p. 1125-9.
19. Yamamoto, J., et al., *Expression and characterization of recombinant mouse mastocytoma histidine decarboxylase.* Biochim Biophys Acta, 1993. **1216**(3): p. 431-40.
20. Schneider, E., et al., *Histamine-producing cell-stimulating activity. Interleukin 3 and granulocyte-macrophage colony-stimulating factor induce de novo synthesis of histidine decarboxylase in hemopoietic progenitor cells.* J Immunol, 1987. **139**(11): p. 3710-7.

21. Coton, E., G.C. Rollan, and A. Lonvaud-Funel, *Histidine carboxylase of Leuconostoc oenos 9204: purification, kinetic properties, cloning and nucleotide sequence of the hdc gene.* J Appl Microbiol, 1998. **84**(2): p. 143-51.
22. Percudani, R. and A. Peracchi, *A genomic overview of pyridoxal-phosphate-dependent enzymes.* EMBO reports, 2003. **4**(9): p. 850-854.
23. Eliot, A.C. and J.F. Kirsch, *Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations.* Annual Review of Biochemistry, 2004. **73**(1): p. 383-415.
24. Jansonius, J.N., *Structure, evolution and action of vitamin B6-dependent enzymes.* Current Opinion in Structural Biology, 1998. **8**(6): p. 759-769.
25. Sandmeier, E., T.I. Hale, and P. Christen, *Multiple evolutionary origin of pyridoxal-5'-phosphate-dependent amino acid decarboxylases.* European Journal of Biochemistry, 1994. **221**(3): p. 997-1002.
26. Moya-Garcia, A.A., M.A. Medina, and F. Sanchez-Jimenez, *Mammalian histidine decarboxylase: from structure to function.* Bioessays, 2005. **27**(1): p. 57-63.
27. Levine, R.J. and W.W. Noll, *Histidine decarboxylase and its inhibition.* Ann N Y Acad Sci, 1969. **166**(1): p. 246-56.
28. Giardina, G., et al., *Open conformation of human DOPA decarboxylase reveals the mechanism of PLP addition to Group II decarboxylases.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(51): p. 20514-20519.
29. Burkhard, P., et al., *Structural insight into Parkinson's disease treatment from drug-inhibited DOPA decarboxylase.* Nat Struct Biol, 2001. **8**(11): p. 963-7.
30. Langendorf, C.G., et al., *Structural characterization of the mechanism through which human glutamic acid decarboxylase auto-activates.* Biosci Rep, 2013. **33**(1): p. 137-44.
31. Fenalti, G., et al., *GABA production by glutamic acid decarboxylase is regulated by a dynamic catalytic loop.* Nat Struct Mol Biol, 2007. **14**(4): p. 280-6.
32. Yatsunami, K., et al., *Structure of the L-histidine decarboxylase gene.* J Biol Chem, 1994. **269**(2): p. 1554-9.
33. Zahnow, C.A., et al., *Cloning of the cDNA encoding human histidine decarboxylase from an erythroleukemia cell line and mapping of the gene locus to chromosome 15.* DNA Seq, 1991. **1**(6): p. 395-400.
34. Yamamoto, M., et al., *Activity and tissue-specific expression of the transcription factor NF-E1 multigene family.* Genes Dev, 1990. **4**(10): p. 1650-62.
35. Orkin, S.H., *Globin gene regulation and switching: circa 1990.* Cell, 1990. **63**(4): p. 665-72.
36. Zon, L.I., et al., *GATA-binding transcription factors in mast cells regulate the promoter of the mast cell carboxypeptidase A gene.* J Biol Chem, 1991. **266**(34): p. 22948-53.
37. Zhang, Z., et al., *The Human Histidine Decarboxylase Promoter Is Regulated by Gastrin and Phorbol 12-Myristate 13-Acetate through a Downstream cis-Acting Element.* Journal of Biological Chemistry, 1996. **271**(24): p. 14188-14197.
38. Hocker, M., et al., *Oxidative stress activates the human histidine decarboxylase promoter in AGS gastric cancer cells.* J Biol Chem, 1998. **273**(36): p. 23046-54.
39. Pacilio, M., et al., *Thrombopoietin induces histidine decarboxylase gene expression in c-mpl transfected UT7 cells.* Biochem Biophys Res Commun, 2001. **285**(5): p. 1095-101.
40. Wessler, S., et al., *Helicobacter pylori activates the histidine decarboxylase promoter through a mitogen-activated protein kinase pathway independent of pathogenicity island-encoded virulence factors.* J Biol Chem, 2000. **275**(5): p. 3629-36.
41. Yu, X., et al., *Circadian factor BMAL1 in histaminergic neurons regulates sleep architecture.* Curr Biol, 2014. **24**(23): p. 2838-44.

42. Shan, L., et al., *Diurnal fluctuation in histidine decarboxylase expression, the rate limiting enzyme for histamine production, and its disorder in neurodegenerative diseases.* Sleep, 2012. **35**(5): p. 713-5.
43. Engel, N., et al., *Experimental evidence for structure-activity features in common between mammalian histidine decarboxylase and ornithine decarboxylase.* Biochem J, 1996. **320 ( Pt 2)**: p. 365-8.
44. Viguera, E., et al., *Mammalian L-amino acid decarboxylases producing 1,4-diamines: analogies among differences.* Trends Biochem Sci, 1994. **19**(8): p. 318-9.
45. Rechsteiner, M. and S.W. Rogers, *PEST sequences and regulation by proteolysis.* Trends Biochem Sci, 1996. **21**(7): p. 267-71.
46. Rogers, S., R. Wells, and M. Rechsteiner, *Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis.* Science, 1986. **234**(4774): p. 364-368.
47. Yatsunami, K., et al., *Comparative studies of human recombinant 74- and 54-kDa L-histidine decarboxylases.* J Biol Chem, 1995. **270**(51): p. 30813-7.
48. Tanaka, S., et al., *Degradation of the 74 kDa form of L-histidine decarboxylase via the ubiquitin-proteasome pathway in a rat basophilic/mast cell line (RBL-2H3).* FEBS Lett, 1997. **417**(2): p. 203-7.
49. Tanaka, S. and A. Ichikawa, *Intracellular localization of histidine decarboxylase.* Inflamm Res, 2001. **50 Suppl 2**: p. S98-9.
50. Komori, H., et al., *Purification, crystallization and preliminary X-ray analysis of human histidine decarboxylase.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2012. **68**(Pt 6): p. 675-7.
51. Cerqueira, N.M., P.A. Fernandes, and M.J. Ramos, *Computational Mechanistic Studies Addressed to the Transimination Reaction Present in All Pyridoxal 5'-Phosphate-Requiring Enzymes.* J Chem Theory Comput, 2011. **7**(5): p. 1356-68.
52. Oliveira, E.F., et al., *Mechanism of formation of the internal aldimine in pyridoxal 5'-phosphate-dependent enzymes.* J Am Chem Soc, 2011. **133**(39): p. 15496-505.
53. Snell, E.E., *Vitamin B6 and decarboxylation of histidine.* Ann N Y Acad Sci, 1990. **585**: p. 1-12.
54. Olmo, M.T., et al., *Spectroscopic analysis of recombinant rat histidine decarboxylase.* J Biochem, 2002. **132**(3): p. 433-9.
55. Hayashi, H., H. Mizuguchi, and H. Kagamiyama, *Rat liver aromatic L-amino acid decarboxylase: spectroscopic and kinetic analysis of the coenzyme and reaction intermediates.* Biochemistry, 1993. **32**(3): p. 812-8.
56. Rodriguez-Caso, C., et al., *Local changes in the catalytic site of mammalian histidine decarboxylase can affect its global conformation and stability.* Eur J Biochem, 2003. **270**(21): p. 4376-87.
57. Battersby, A.R., et al., *Studies of enzyme-mediated reactions. Part 13. Stereochemical course of the formation of histamine by decarboxylation of (2S)-histidine with enzymes from Clostridium welchii and Lactobacillus 30a.* J Chem Soc Perkin 1, 1980. **1**: p. 43-51.
58. Battersby, A.R., R. Joyeau, and J. Staunton, *Stereochemical study of formation of histamine from (2S)-histidine by mammalian histidine decarboxylase.* FEBS Lett, 1979. **107**(1): p. 231-2.
59. Hayashi, H., et al., *Recent topics in pyridoxal 5'-phosphate enzyme studies.* Annu Rev Biochem, 1990. **59**: p. 87-110.
60. Dunathan, H.C., *Conformation and reaction specificity in pyridoxal phosphate enzymes.* Proc Natl Acad Sci U S A, 1966. **55**(4): p. 712-6.
61. Dunathan, H.C., *Stereochemical aspects of pyridoxal phosphate catalysis.* Adv Enzymol Relat Areas Mol Biol, 1971. **35**: p. 79-134.

62. Tahanejad, F.S. and H. Naderi-Manesh, *Quantum mechanical study of the intermediates formed following the reaction of the histidine decarboxylase's substrate and inhibitors with coenzyme.* Eur J Med Chem, 2000. **35**(3): p. 283-9.
63. Malcolm, B.A. and J.F. Kirsch, *Site-directed mutagenesis of aspartate aminotransferase from E. coli.* Biochem Biophys Res Commun, 1985. **132**(3): p. 915-21.
64. Kuramitsu, S., et al., *Substitution of an arginyl residue for the active site lysyl residue (Lys258) of aspartate aminotransferase.* Biochem Biophys Res Commun, 1987. **146**(2): p. 416-21.
65. Vaaler, G.L. and E.E. Snell, *Pyridoxal 5'-phosphate dependent histidine decarboxylase: overproduction, purification, biosynthesis of soluble site-directed mutant proteins, and replacement of conserved residues.* Biochemistry, 1989. **28**(18): p. 7306-13.
66. Fleming, J.V., et al., *Mapping of catalytically important residues in the rat L-histidine decarboxylase enzyme using bioinformatic and site-directed mutagenesis approaches.* Biochem J, 2004. **379**(Pt 2): p. 253-61.
67. Pino-Angeles, A., et al., *Substrate uptake and protein stability relationship in mammalian histidine decarboxylase.* Proteins, 2010. **78**(1): p. 154-61.
68. Ishii, S., et al., *Aromatic L-amino acid decarboxylase: conformational change in the flexible region around Arg334 is required during the transaldimination process.* Protein Sci, 1998. **7**(8): p. 1802-10.
69. Moya-Garcia, A.A., et al., *Analysis of the decarboxylation step in mammalian histidine decarboxylase. A computational study.* J Biol Chem, 2008. **283**(18): p. 12393-401.
70. Caro-Astorga, J., et al., *Nascent histamine induces alpha-synuclein and caspase-3 on human cells.* Biochem Biophys Res Commun, 2014. **451**(4): p. 580-6.
71. Pagotto, R.M., et al., *Histamine inhibits adrenocortical cell proliferation but does not affect steroidogenesis.* J Endocrinol, 2014. **221**(1): p. 15-28.
72. Niijima-Yaoita, F., et al., *Roles of histamine in exercise-induced fatigue: favouring endurance and protecting against exhaustion.* Biol Pharm Bull, 2012. **35**(1): p. 91-7.
73. Na, J.I., et al., *Histidine decarboxylase expression influences the neofolliculogenesis of newborn mouse dermal cells.* J Dermatol Sci, 2012. **67**(2): p. 95-100.
74. Nakazawa, S., et al., *Histamine synthesis is required for granule maturation in murine mast cells.* Eur J Immunol, 2014. **44**(1): p. 204-14.
75. Ohtsu, H., et al., *Mice lacking histidine decarboxylase exhibit abnormal mast cells.* FEBS Lett, 2001. **502**(1-2): p. 53-6.
76. Eichenfield, L.F., et al., *Guidelines of care for the management of atopic dermatitis: section 2. Management and treatment of atopic dermatitis with topical therapies.* J Am Acad Dermatol, 2014. **71**(1): p. 116-32.
77. Sidbury, R., et al., *Guidelines of care for the management of atopic dermatitis: section 3. Management and treatment with phototherapy and systemic agents.* J Am Acad Dermatol, 2014. **71**(2): p. 327-49.
78. Danby, S.G., *A new perspective on histamine in atopic dermatitis.* Br J Dermatol, 2014. **171**(4): p. 688.
79. Gutowska-Owsiak, D., et al., *The histamine-synthesizing enzyme histidine decarboxylase is upregulated by keratinocytes in atopic skin.* Br J Dermatol, 2014. **171**(4): p. 771-8.
80. Gutowska-Owsiak, D., et al., *Histamine enhances keratinocyte-mediated resolution of inflammation by promoting wound healing and response to infection.* Clin Exp Dermatol, 2014. **39**(2): p. 187-95.

81. Tamaka, K., et al., *Histamine suppresses regulatory T cells mediated by TGF-beta in murine chronic allergic contact dermatitis.* Exp Dermatol, 2015. **24**(4): p. 280-4.
82. Gervasini, G., et al., *Variability of the L-Histidine decarboxylase gene in allergic rhinitis.* Allergy, 2010. **65**(12): p. 1576-84.
83. Guth, P.H. and P. Hall, *Microcirculatory and mast cell changes in restraint-induced gastric ulcer.* Gastroenterology, 1966. **50**(4): p. 562-70.
84. Levine, R.J. and E.C. Senay, *Histamine in the pathogenesis of stress ulcers in the rat.* Am J Physiol, 1968. **214**(4): p. 892-6.
85. Parmar, N.S., G. Hennings, and O.P. Gulati, *Histidine decarboxylase inhibition: a novel approach towards the development of an effective and safe gastric anti-ulcer drug.* Agents Actions, 1984. **15**(5-6): p. 494-9.
86. Chen, D., et al., *Acute responses of rat stomach enterochromaffinlike cells to gastrin: secretory activation and adaptation.* Gastroenterology, 1994. **107**(1): p. 18-27.
87. Sandvik, A.K., et al., *Gastrin regulates histidine decarboxylase activity and mRNA abundance in rat oxyntic mucosa.* Am J Physiol, 1994. **267**(2 Pt 1): p. G254-8.
88. Ku, H.J., et al., *Bile acid increases expression of the histamine-producing enzyme, histidine decarboxylase, in gastric cells.* World J Gastroenterol, 2014. **20**(1): p. 175-82.
89. Alkan, M., et al., *Histidine Decarboxylase Deficiency Prevents Autoimmune Diabetes in NOD Mice.* J Diabetes Res, 2015. **2015**: p. 965056.
90. Zhang, F., et al., *HDC gene polymorphisms are associated with age at natural menopause in Caucasian women.* Biochemical and Biophysical Research Communications, 2006. **348**(4): p. 1378-1382.
91. Mackins, C.J., et al., *Cardiac mast cell-derived renin promotes local angiotensin formation, norepinephrine release, and arrhythmias in ischemia/reperfusion.* J Clin Invest, 2006. **116**(4): p. 1063-70.
92. Kim, J., et al., *Impact of blockade of histamine H2 receptors on chronic heart failure revealed by retrospective and prospective randomized studies.* J Am Coll Cardiol, 2006. **48**(7): p. 1378-84.
93. He, G.H., et al., *Relation of polymorphism of the histidine decarboxylase gene to chronic heart failure in Han Chinese.* Am J Cardiol, 2015. **115**(11): p. 1555-62.
94. Matsuki, Y., et al., *Histidine decarboxylase expression as a new sensitive and specific marker for small cell lung carcinoma.* Mod Pathol, 2003. **16**(1): p. 72-8.
95. Tanimoto, A., et al., *Histidine decarboxylase expression in pancreatic endocrine cells and related tumors.* Pathol Int, 2004. **54**(6): p. 408-12.
96. Boer, K., et al., *Expression of histidine decarboxylase in human colonic cancer cells and adenomatous polyps.* Inflamm Res, 2003. **52 Suppl 1**: p. S76-7.
97. Darvas, Z., et al., *Autonomous histamine metabolism in human melanoma cells.* Melanoma Res, 2003. **13**(3): p. 239-46.
98. He, G.H., et al., *Associations of polymorphisms in histidine decarboxylase, histamine N-methyltransferase and histamine receptor H3 genes with breast cancer.* PLoS One, 2014. **9**(5): p. e97728.
99. Cui, J., et al., *The expression of non-mast histamine in tumor associated microvessels in human colorectal cancers.* Pathol Oncol Res, 2013. **19**(2): p. 311-6.
100. Francis, H., et al., *Inhibition of histidine decarboxylase ablates the autocrine tumorigenic effects of histamine in human cholangiocarcinoma.* Gut, 2012. **61**(5): p. 753-64.
101. Levine, R.J. and D.E. Watts, *Inhibition of histidine decarboxylase activity in vitro by norepinephrine and related compounds.* Biochem Pharmacol, 1967. **16**(6): p. 993-1000.

102. Levine, R.J., *Histamine synthesis in man: inhibition by 4-bromo-3-hydroxybenzyloxyamine.* Science, 1966. **154**(3752): p. 1017-9.
103. Kuntzman, R., et al., *Inhibition of norepinephrine synthesis in mouse brain by blockade of dopaminebeta-oxidase.* Life Sci, 1962. **1**: p. 85-92.
104. Free, C.A., E. Majchrowicz, and S.M. Hess, *Mechanism of inhibition of histidine decarboxylase by rhodanines.* Biochem Pharmacol, 1971. **20**(7): p. 1421-8.
105. Taylor, R.J., Jr., *Inhibition of gastric acid secretion by 4-imidazolyl-3-amino-2-butanone (McN-A-1293), a specific inhibitor of histidine decarboxylase.* Biochem Pharmacol, 1978. **27**(22): p. 2653-4.
106. Taylor, R.J., Jr., F.J. Leinweber, and G.A. Braun, *4-imidazolyl-3-amino-2-butanone (McN-A-1293), a new specific inhibitor of histidine decarboxylase.* Biochem Pharmacol, 1973. **22**(18): p. 2299-310.
107. Kelley, J.L., C.A. Miller, and H.L. White, *Inhibition of histidine decarboxylase. Derivatives of histidine.* J Med Chem, 1977. **20**(4): p. 506-9.
108. Hammar, L. and U. Ragnarsson, *Peptide inhibition of mammalian histidine decarboxylase.* Agents Actions, 1979. **9**(4): p. 314-8.
109. Awapara, J., R.P. Sandman, and C. Hanly, *Activation of DOPA decarboxylase by pyridoxal phosphate.* Arch Biochem Biophys, 1962. **98**: p. 520-5.
110. Kollonitsch, J., et al., *Selective inhibitors of biosynthesis of aminergic neurotransmitters.* Nature, 1978. **274**(5674): p. 906-8.
111. Kubota, H., et al., *Mechanism of inactivation of mammalian L-histidine decarboxylase by (S)-alpha-fluoromethylhistidine.* Biochem Pharmacol, 1984. **33**(7): p. 983-90.
112. Tanase, S., B.M. Guirard, and E.E. Snell, *Purification and properties of a pyridoxal 5'-phosphate-dependent histidine decarboxylase from Morganella morganii AM-15.* J Biol Chem, 1985. **260**(11): p. 6738-46.
113. Hayashi, H., S. Tanase, and E.E. Snell, *Pyridoxal 5'-phosphate-dependent histidine decarboxylase. Inactivation by alpha-fluoromethylhistidine and comparative sequences at the inhibitor- and coenzyme-binding sites.* J Biol Chem, 1986. **261**(24): p. 11003-9.
114. Recsei, P.A. and E.E. Snell, *Pyruvoyl enzymes.* Annu Rev Biochem, 1984. **53**: p. 357-87.
115. Wada, H., et al., *Mammalian histidine decarboxylase and its suicide substrate alpha-fluoromethylhistidine.* Prog Clin Biol Res, 1984. **144A**: p. 245-54.
116. Campos, H.A., et al., *Alpha-fluoromethylhistidine, an inhibitor of histamine biosynthesis, causes arterial hypertension.* Naunyn Schmiedebergs Arch Pharmacol, 1996. **354**(5): p. 627-32.
117. Cabrera, C., R. Artacho, and R. Gimenez, *Beneficial effects of green tea--a review.* J Am Coll Nutr, 2006. **25**(2): p. 79-99.
118. Melgarejo, E., et al., *(-)-Epigallocatechin-3-gallate interferes with mast cell adhesiveness, migration and its potential to recruit monocytes.* Cell Mol Life Sci, 2007. **64**(19-20): p. 2690-701.
119. Melgarejo, E., et al., *Epigallocatechin gallate reduces human monocyte mobility and adhesion in vitro.* Br J Pharmacol, 2009. **158**(7): p. 1705-12.
120. Bachrach, U. and Y.C. Wang, *Cancer therapy and prevention by green tea: role of ornithine decarboxylase.* Amino Acids, 2002. **22**(1): p. 1-13.
121. Tosetti, F., et al., *Angioprevention': angiogenesis is a common and key target for cancer chemopreventive agents.* FASEB J, 2002. **16**(1): p. 2-14.
122. Ho, Y.C., et al., *Epigallocatechin-3-gallate inhibits the invasion of human oral cancer cells and decreases the productions of matrix metalloproteinases and urokinase-plasminogen activator.* J Oral Pathol Med, 2007. **36**(10): p. 588-93.
123. Albrecht, D.S., et al., *Epigallocatechin-3-gallate (EGCG) inhibits PC-3 prostate cancer cell proliferation via MEK-independent ERK1/2 activation.* Chem Biol Interact, 2008. **171**(1): p. 89-95.

124. Rodriguez-Caso, C., et al., *Green tea epigallocatechin-3-gallate is an inhibitor of mammalian histidine decarboxylase.* Cell Mol Life Sci, 2003. **60**(8): p. 1760-3.
125. Bertoldi, M., M. Gonsalvi, and C.B. Voltattorni, *Green tea polyphenols: novel irreversible inhibitors of dopa decarboxylase.* Biochem Biophys Res Commun, 2001. **284**(1): p. 90-3.
126. Ruiz-Perez, M.V., et al., *Structural perspective on the direct inhibition mechanism of EGCG on mammalian histidine decarboxylase and DOPA decarboxylase.* J Chem Inf Model, 2012. **52**(1): p. 113-9.
127. Bakrania, A.K. and S.S. Patel, *Combination treatment for allergic conjunctivitis - Plant derived histidine decarboxylase inhibitor and H1 antihistaminic drug.* Exp Eye Res, 2015. **137**: p. 32-8.
128. Mishiro, Y., et al., *Use of angiotensin II stress pulsed tissue Doppler imaging to evaluate regional left ventricular contractility in patients with hypertrophic cardiomyopathy.* J Am Soc Echocardiogr, 2000. **13**(12): p. 1065-73.
129. Rogoza, R.M., et al., *Electron spin resonance spectroscopy reveals alpha-phenyl-N-tert-butylnitrone spin-traps free radicals in rat striatum and prevents haloperidol-induced vacuous chewing movements in the rat model of human tardive dyskinesia.* Synapse, 2004. **54**(3): p. 156-63.
130. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: Survey of methods for simulating the activity of proteins.* Chemical Reviews, 2006. **106**(5): p. 1589-1615.
131. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nat Struct Biol, 2002. **9**(9): p. 646-52.
132. Vanommeslaeghe, K., O. Guvench, and A.D. MacKerell, *Molecular Mechanics.* Current Pharmaceutical Design, 2014. **20**(20): p. 3281-3292.
133. Leach, A., *Molecular Modelling: Principles and Applications (2nd Edition)*. 2001: Prentice Hall.
134. Case, D.A., et al., *AMBER 12*. 2012, University of California, San Francisco.
135. Hornak, V., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins, 2006. **65**(3): p. 712-25.
136. Brooks, B.R., et al., *Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
137. Jorgensen, W.L. and J. Tiradorives, *The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin.* Journal of the American Chemical Society, 1988. **110**(6): p. 1657-1666.
138. Christen, M., et al., *The GROMOS software for biomolecular simulation: GROMOS05.* Journal of Computational Chemistry, 2005. **26**(16): p. 1719-1751.
139. Alder, B.J. and T.E. Wainwright, *Studies in Molecular Dynamics .1. General Method.* Journal of Chemical Physics, 1959. **31**(2): p. 459-466.
140. Planck, M., *Ueber die Elementarquanta der Materie und der Elektricität.* Annalen der Physik, 1901. **309**(3): p. 564-566.
141. Jinxin, X., et al., *A determination of the Planck constant by the generalized joule balance method with a permanent-magnet system at NIM.* Metrologia, 2016. **53**(1): p. 86.
142. Einstein, A., *Generation and conversion of light with regard to a heuristic point of view.* Annalen Der Physik, 1905. **17**(6): p. 132-148.
143. Hertz, H., *Ueber einen Einfluss des ultravioletten Lichtes auf die electrische Entladung.* Annalen der Physik, 1887. **267**(8): p. 983-1000.
144. Compton, A.H., *A Quantum Theory of the Scattering of X-rays by Light Elements.* Physical Review, 1923. **21**(5): p. 483-502.
145. Bohr, N., *On the Constitution of Atoms and Molecules.* Philosophical Magazine, 1913. **26**(155): p. 857-875.

146. Heisenberg, W., *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik.* Zeitschrift für Physik. **43**(3): p. 172-198.
147. Dorobantu, V., *The Postulates of Quantum Mechanics.* Quantum Computability, 2006. **1**: p. 1-26.
148. Nottale, L. and M.N. Celerier, *Derivation of the postulates of quantum mechanics from the first principles of scale relativity.* Journal of Physics a-Mathematical and Theoretical, 2007. **40**(48): p. 14471-14498.
149. Cramer, C.J., *Essentials of Computational Chemistry - Theories and Models*. 2nd ed. 2004: John Wiley & Sons. 596.
150. Born, M. and R. Oppenheimer, *Zur Quantentheorie der Molekeln.* Annalen der Physik, 1927. **389**(20): p. 457-484.
151. Jensen, F., *Introduction to Computational Chemistry*. 2nd ed. 2007: John Wiley & Sons.
152. Atkins, P. and R. Friedman, *Molecular Quantum Mechanics*. 4th ed. 2005: OXFORD UNIVERSITY PRESS. 573.
153. Stewart, J.J.P., *Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements.* Journal of Molecular Modeling, 2007. **13**(12): p. 1173-1213.
154. Thomas, L.H., *The calculation of atomic fields.* Mathematical Proceedings of the Cambridge Philosophical Society, 1927. **23**(05): p. 542-548.
155. Fermi, E., *Un metodo statistico per la determinazione di alcune priorietà dell'atome.* Rend. Accad. Naz. Lincei, 1927. **6**: p. 602–607.
156. Hohenberg, P. and W. Kohn, *Inhomogeneous Electron Gas.* Physical Review, 1964. **136**(3B): p. B864-B871.
157. Kohn, W. and L.J. Sham, *Self-Consistent Equations Including Exchange and Correlation Effects.* Physical Review, 1965. **140**(4A): p. A1133-A1138.
158. Yanai, T., D.P. Tew, and N.C. Handy, *A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP).* Chemical Physics Letters, 2004. **393**(1–3): p. 51-57.
159. Filatov, M. and W. Thiel, *A new gradient-corrected exchange-correlation density functional.* Molecular Physics, 1997. **91**(5): p. 847-860.
160. Perdew, J.P. and W. Yue, *Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation.* Physical Review B, 1986. **33**(12): p. 8800-8802.
161. Becke, A.D., *Density-functional exchange-energy approximation with correct asymptotic behavior.* Physical Review A, 1988. **38**(6): p. 3098-3100.
162. Lacks, D.J. and R.G. Gordon, *Pair interactions of rare-gas atoms as a test of exchange-energy-density functionals in regions of large density gradients.* Physical Review A, 1993. **47**(6): p. 4681-4690.
163. Paier, J., et al., *The Perdew–Burke–Ernzerhof exchange-correlation functional applied to the G2-1 test set using a plane-wave basis set.* The Journal of Chemical Physics, 2005. **122**(23): p. 234102.
164. Perdew, J.P., *Generalized gradient approximations for exchange and correlation: A look backward and forward.* Physica B: Condensed Matter, 1991. **172**(1–2): p. 1-6.
165. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B, 1988. **37**(2): p. 785-789.
166. Zhao, Y., N.E. Schultz, and D.G. Truhlar, *Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions.* The Journal of Chemical Physics, 2005. **123**(16): p. 161103.
167. Zhao, Y., N.E. Schultz, and D.G. Truhlar, *Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for*

*Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions.* Journal of Chemical Theory and Computation, 2006. **2**(2): p. 364-382.
168. Zhao, Y. and D.G. Truhlar, *The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals.* Theoretical Chemistry Accounts, 2007. **120**(1): p. 215-241.
169. Grimme, S., et al., *A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu.* The Journal of Chemical Physics, 2010. **132**(15): p. 154104.
170. Grimme, S., S. Ehrlich, and L. Goerigk, *Effect of the damping function in dispersion corrected density functional theory.* J Comput Chem, 2011. **32**(7): p. 1456-65.
171. Young, D., *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*. 2001: John Wiley & Sons.
172. Dapprich, S., et al., *A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives.* Journal of Molecular Structure-Theochem, 1999. **461**: p. 1-21.
173. Wang, J., et al., *Development and testing of a general amber force field.* J Comput Chem, 2004. **25**(9): p. 1157-74.
174. Mahoney, M.W. and W.L. Jorgensen, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions.* The Journal of Chemical Physics, 2000. **112**(20): p. 8910-8922.
175. Larini, L., R. Mannella, and D. Leporini, *Langevin stabilization of molecular-dynamics simulations of polymers by means of quasisymplectic algorithms.* J Chem Phys, 2007. **126**(10): p. 104101.
176. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
177. Roe, D.R. and T.E. Cheatham, 3rd, *PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.* J Chem Theory Comput, 2013. **9**(7): p. 3084-95.
178. Frisch, M.J., et al., *Gaussian 09*. 2009, Gaussian, Inc.: Wallingford, CT, USA.
179. Roy Dennington, T.K., John Millam, *GaussView*. 2009, Semichem Inc.: Shawnee Mission.

This page was intentionally left blank.

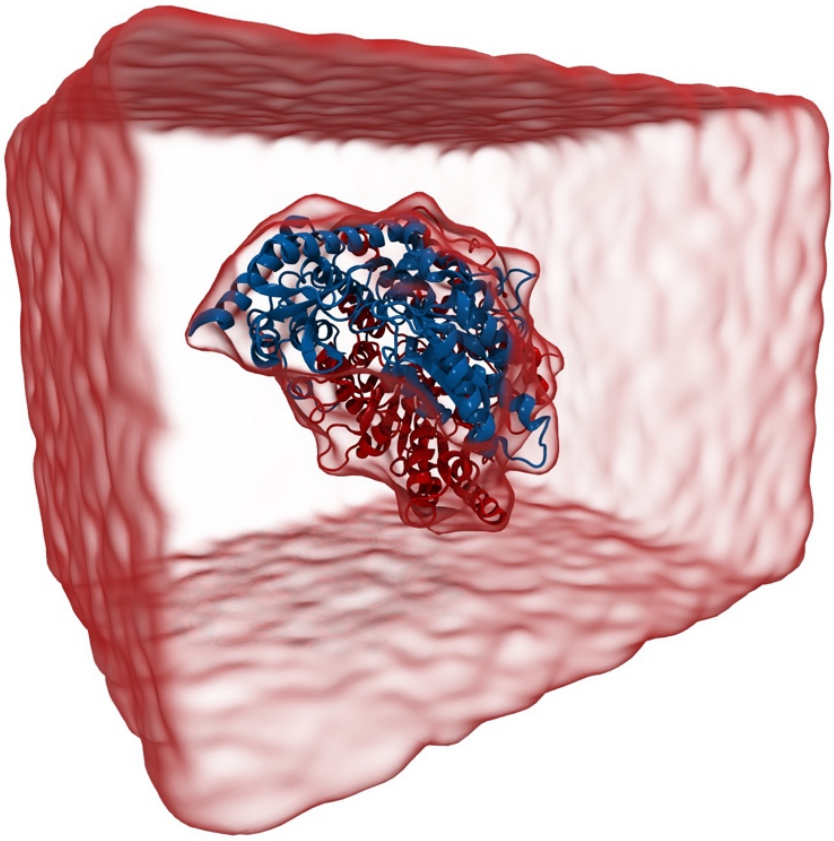
# G. Appendices

This page was intentionally left blank.

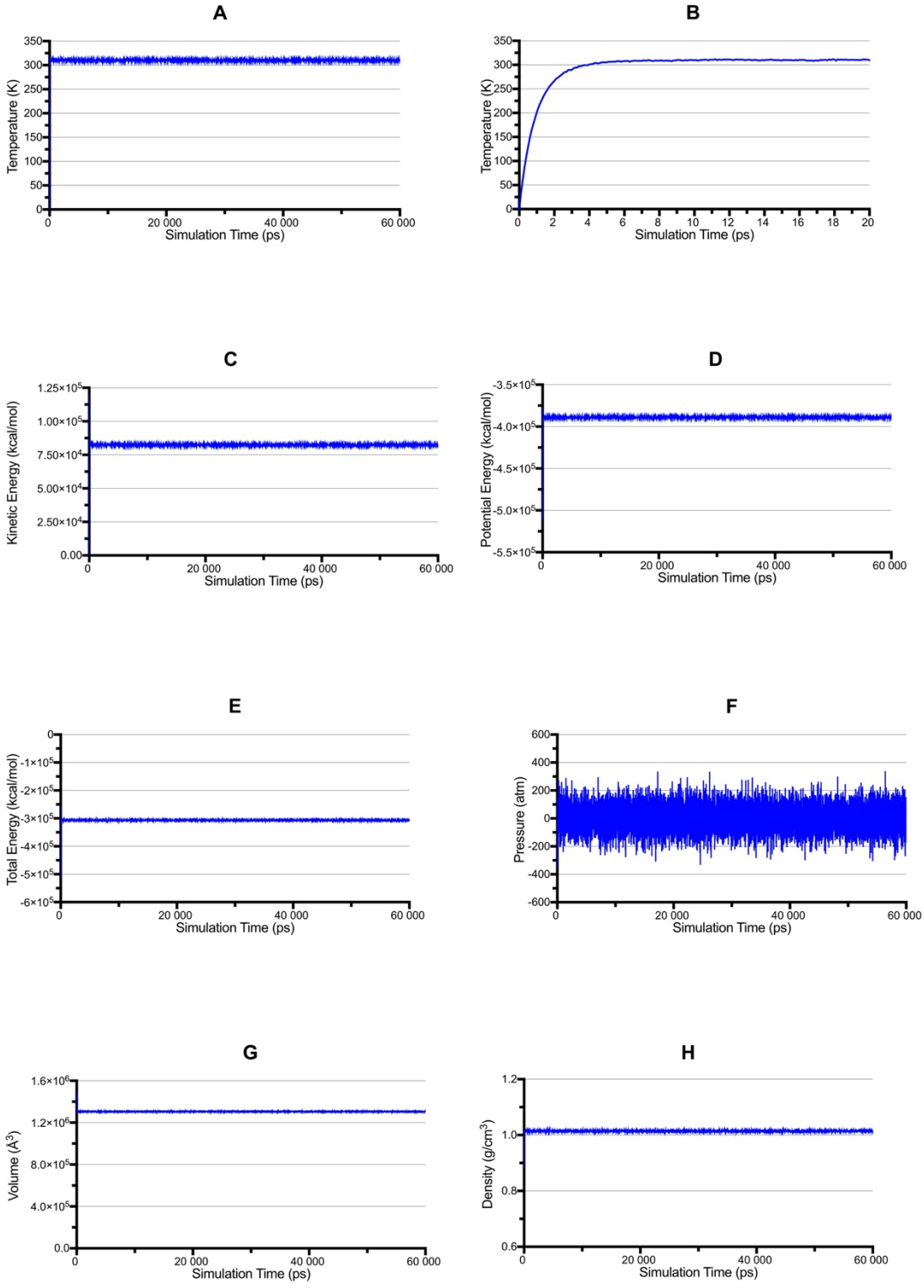
# 1. Detailed information about the molecular dynamics simulations

```
CHAIN A:     -SMEPEEYRERGREMVDYICQYLSTVRERRVTPDVQPGYLRAQLPESAPEDPDSWDSIFGDIERIIM
CHAIN B:     -SMEPEEYRERGREMVDYICQYLSTVRERRVTPDVQPGYLRAQLPESAPEDPDSWDSIFGDIERIIM
CHAIN C: GPLGSMEPEEYRERGREMVDYICQYLSTVRERRVTPDVQPGYLRAQLPESAPEDPDSWDSIFGDIERIIM
CHAIN D:    GSMEPEEYRERGREMVDYICQYLSTVRERRVTPDVQPGYLRAQLPESAPEDPDSWDSIFGDIERIIM
CHAIN E:    GSMEPEEYRERGREMVDYICQYLSTVRERRVTPDVQPGYLRAQLPESAPEDPDSWDSIFGDIERIIM
CHAIN F:    GSMEPEEYRERGREMVDYICQYLSTVRERRVTPDVQPGYLRAQLPESAPEDPDSWDSIFGDIERIIM
            70        80        90        100       110       120       130
CHAIN A: PGVVHWQSPHMHAYYPALTSWPSLLGDMLADAINCLGFTWASSPACTELEMNVMDWLAKMLGLPEHFLHH
CHAIN B: PGVVHWQSPHMHAYYPALTSWPSLLGDMLADAINCLGFTWASSPACTELEMNVMDWLAKMLGLPEHFLHH
CHAIN C: PGVVHWQSPHMHAYYPALTSWPSLLGDMLADAINCLGFTWASSPACTELEMNVMDWLAKMLGLPEHFLHH
CHAIN D: PGVVHWQSPHMHAYYPALTSWPSLLGDMLADAINCLGFTWASSPACTELEMNVMDWLAKMLGLPEHFLHH
CHAIN E: PGVVHWQSPHMHAYYPALTSWPSLLGDMLADAINCLGFTWASSPACTELEMNVMDWLAKMLGLPEHFLHH
CHAIN F: PGVVHWQSPHMHAYYPALTSWPSLLGDMLADAINCLGFTWASSPACTELEMNVMDWLAKMLGLPEHFLHH
            140       150       160       170       180       190       200
CHAIN A: HPSSQGGGVLQSTVSESTLIALLAARKNKILEMKTSEPDADESSLNARLVAYASDQAHSSVEKAGLISLV
CHAIN B: HPSSQGGGVLQSTVSESTLIALLAARKNKILEMKTSEPDADESSLNARLVAYASDQAHSSVEKAGLISLV
CHAIN C: HPSSQGGGVLQSTVSESTLIALLAARKNKILEMKTSEPDADESSLNARLVAYASDQAHSSVEKAGLISLV
CHAIN D: HPSSQGGGVLQSTVSESTLIALLAARKNKILEMKTSEPDADESSLNARLVAYASDQAHSSVEKAGLISLV
CHAIN E: HPSSQGGGVLQSTVSESTLIALLAARKNKILEMKTSEPDADESSLNARLVAYASDQAHSSVEKAGLISLV
CHAIN F: HPSSQGGGVLQSTVSESTLIALLAARKNKILEMKTSEPDADESSLNARLVAYASDQAHSSVEKAGLISLV
            210       220       230       240       250       260       270
CHAIN A: KMKFLPVDDNFSLRGEALQKAIEEDKQRGLVPVFVCATLGTTGVCAFD?LSELGPICAREGLWLHIDAAY
CHAIN B: KMKFLPVDDNFSLRGEALQKAIEEDKQRGLVPVFVCATLGTTGVCAFD?LSELGPICAREGLWLHIDAAY
CHAIN C: KMKFLPVDDNFSLRGEALQKAIEEDKQRGLVPVFVCATLGTTGVCAFD?LSELGPICAREGLWLHIDAAY
CHAIN D: KMKFLPVDDNFSLRGEALQKAIEEDKQRGLVPVFVCATLGTTGVCAFD?LSELGPICAREGLWLHIDAAY
CHAIN E: KMKFLPVDDNFSLRGEALQKAIEEDKQRGLVPVFVCATLGTTGVCAFD?LSELGPICAREGLWLHIDAAY
CHAIN F: KMKFLPVDDNFSLRGEALQKAIEEDKQRGLVPVFVCATLGTTGVCAFD?LSELGPICAREGLWLHIDAAY
            280       290       300       310       320       330       340
CHAIN A: AGTAFLCPEFRGFLKGIEYADSFTFNPSKWMMVHFDCTGFWVKDKYKLQQTFSVNPIYLRHANSGVATDF
CHAIN B: AGTAFLCPEFRGFLKGIEYADSFTFNPSKWMMVHFDCTGFWVKDKYKLQQTFSVNPIYLRHANSGVATDF
CHAIN C: AGTAFLCPEFRGFLKGIEYADSFTFNPSKWMMVHFDCTGFWVKDKYKLQQTFSVNPIYLRHANSGVATDF
CHAIN D: AGTAFLCPEFRGFLKGIEYADSFTFNPSKWMMVHFDCTGFWVKDKYKLQQTFSVNPIYLRHANSGVATDF
CHAIN E: AGTAFLCPEFRGFLKGIEYADSFTFNPSKWMMVHFDCTGFWVKDKYKLQQTFSVNPIYLRHANSGVATDF
CHAIN F: AGTAFLCPEFRGFLKGIEYADSFTFNPSKWMMVHFDCTGFWVKDKYKLQQTFSVNPIYLRHANSGVATDF
            350       360       370       380       390       400       410
CHAIN A: MHWQIPLSRRFRSVKLWFVIRSFGVKNLQAHVRHGTEMAKYFESLVRNDPSFEIPAKRHLGLVVFRLKGP
CHAIN B: MHWQIPLSRRFRSVKLWFVIRSFGVKNLQAHVRHGTEMAKYFESLVRNDPSFEIPAKRHLGLVVFRLKGP
CHAIN C: MHWQIPLSRRFRSVKLWFVIRSFGVKNLQAHVRHGTEMAKYFESLVRNDPSFEIPAKRHLGLVVFRLKGP
CHAIN D: MHWQIPLSRRFRSVKLWFVIRSFGVKNLQAHVRHGTEMAKYFESLVRNDPSFEIPAKRHLGLVVFRLKGP
CHAIN E: MHWQIPLSRRFRSVKLWFVIRSFGVKNLQAHVRHGTEMAKYFESLVRNDPSFEIPAKRHLGLVVFRLKGP
CHAIN F: MHWQIPLSRRFRSVKLWFVIRSFGVKNLQAHVRHGTEMAKYFESLVRNDPSFEIPAKRHLGLVVFRLKGP
            420       430       440       450       460       470       480
CHAIN A: NSLTENVLKEIAKAGRLFLIPATIQDKLIIRFTVTSQFTTRDDILRDWNLIRDAATLILSQ---------
CHAIN B: NSLTENVLKEIAKAGRLFLIPATIQDKLIIRFTVTSQFTTRDDILRDWNLIRDAATLILSQ---------
CHAIN C: NSLTENVLKEIAKAGRLFLIPATIQDKLIIRFTVTSQFTTRDDILRDWNLIRDAATLILSQ---------
CHAIN D: NSLTENVLKEIAKAGRLFLIPATIQDKLIIRFTVTSQFTTRDDILRDWNLIRDAATLILSQ---------
CHAIN E: NSLTENVLKEIAKAGRLFLIPATIQDKLIIRFTVTSQFTTRDDILRDWNLIRDAATLILSQ---------
CHAIN F: NSLTENVLKEIAKAGRLFLIPATIQDKLIIRFTVTSQFTTRDDILRDWNLIRDAATLILSQ---------
```

**Supplementary Fig. 1 - Sequence alignment of the amino acid sequence of all chains from PDB file 4E1O. Red boxes highlight a position where an amino acid residue is missing. Yellow boxes highlight the Gly-Pro-Cys tag introduced artificially.**



**Supplementary Fig. 2 - *New* Cartoon representation of the human mHDC protein structure and TIP3P water solvation box. Subunit A and B are colored by blue and red, respectively.**



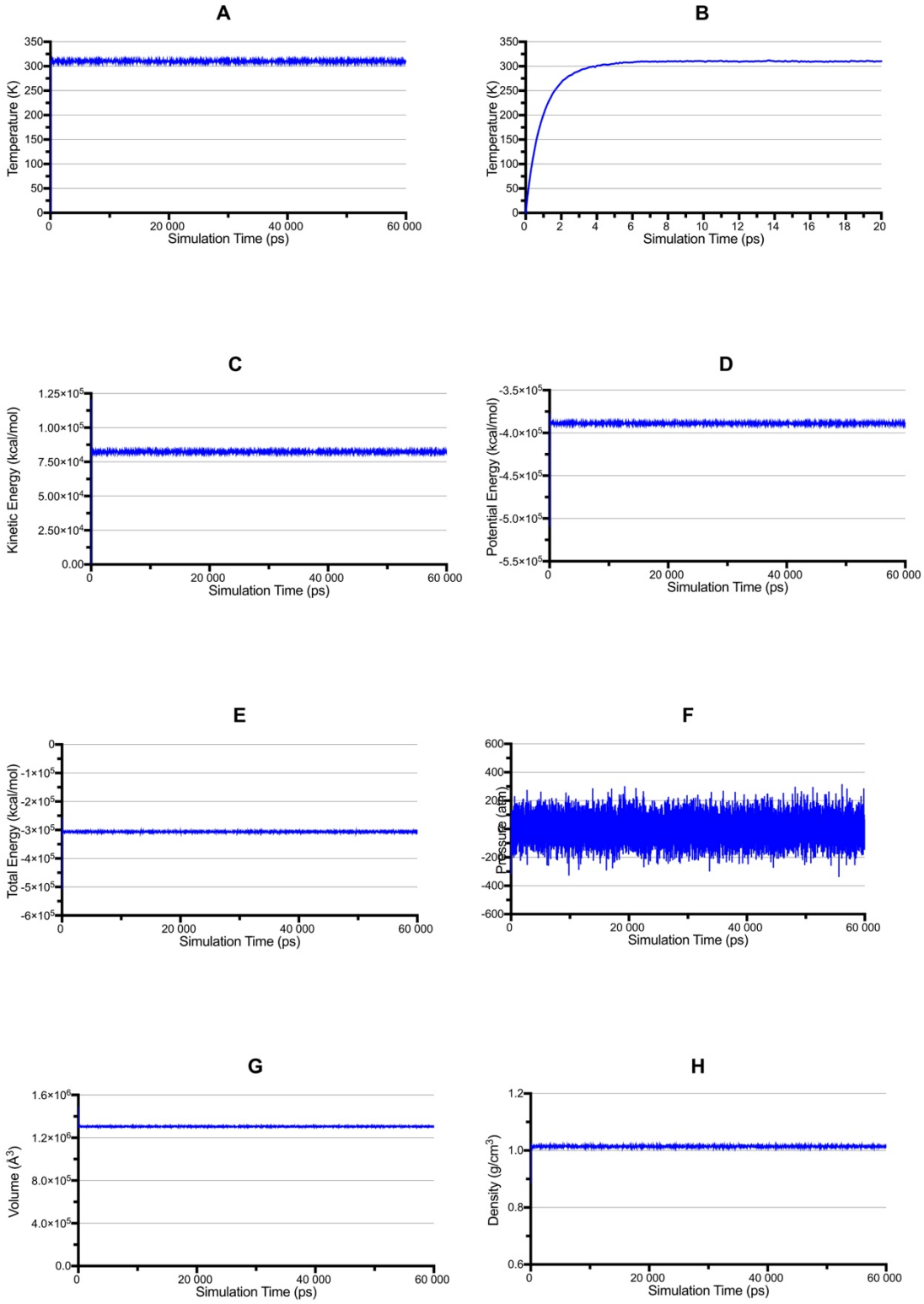

**Supplementary Fig. 3 – Physical general properties during the MD simulation of EA-S model.**
(A) Temperature; (C) Kinetic energy; (D) Potential energy; (E) Total system energy; (F) Pressure; (G) Volume; and (H) Density variations along the 60 ns MD simulation. (B) Temperature variation along the first 20 ps MD simulation corresponding to the equilibration step.

**Supplementary Table 1 - Mean values and respective standard deviations for the physical general properties during MD simulation of EA-S model.**

Calculations were done considering only 20 ns of the MD simulations between 40 and 60 ns, corresponding to an equilibrated region.

| | |
|---|---|
| Temperature (K) | 310.0 ± 5.1 |
| Kinetic Energy (kcal/mol) | $(8.27 \pm 0.36)\times10^{4}$ |
| Total Energy (kcal/mol) | $(-3.06 \pm 0.05)\times10^{5}$ |
| Volume ($Å^3$) | $(1.305 \pm 0.003)\times10^{6}$ |

| | |
|---|---|
| Potential Energy (kcal/mol) | $(-3.89 \pm 0.03)\times10^{5}$ |
| Pressure (atm) | 1.46 ± 85.69 |
| Density ($g/cm^3$) | 1.014 ± 0.002 |




**Supplementary Fig. 4 - Physical general properties during the MD simulation of QI model.**
(A) Temperature; (C) Kinetic energy; (D) Potential energy; (E) Total system energy; (F) Pressure; (G) Volume; and (H) Density variations along the 60 ns MD simulation. (B) Temperature variation along the first 20 ps MD simulation corresponding to the equilibration step.

**Supplementary Table 2 – Mean values and respective standard deviations for the physical general properties during MD simulation of QI model.**

Calculations were done considering only 20 ns of the MD simulations between 40 and 60 ns, corresponding to an equilibrated region.

| | |
|---|---|
| Temperature (K) | 310.0 ± 5.1 |
| Kinetic Energy (kcal/mol) | $(8.27 \pm 0.36)\times10^{4}$ |
| Total Energy (kcal/mol) | $(-3.06 \pm 0.05)\times10^{5}$ |
| Volume ($Å^3$) | $(1.305 \pm 0.003)\times10^{6}$ |

| | |
|---|---|
| Potential Energy (kcal/mol) | $(-3.89 \pm 0.03)\times10^{5}$ |
| Pressure (atm) | 1.31 ± 84.78 |
| Density ($g/cm^3$) | 1.014 ± 0.002 |

## 2. Book chapter

This book chapter was accepted for publication in the multi-volume set (I-V) “Therapeutic Nanostructures” by the Elsevier publisher.

# Cancer therapies based on Enzymatic Amino-Acid Depletion

Henrique Fernandes, Carla Teixeira, P. A. Fernandes, M. J. Ramos and Nuno M. F. S. A. Cerqueira*

UCIBIO@REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal
*Corresponding author (nscerque@fc.up.pt)

Carla Teixeira and Henrique Fernandes contributed equally to this work.

**Keywords:** amino acid deprivation, heterologous enzymes, tumour, cancer, catalytic mechanism, asparaginase, methioninase, arginase, arginine deaminase.

## Abstract

The increasing understanding of tumour biology has allowed the identification of various cellular characteristics that are more frequently associated with cancer cells than with normal cells. These findings have prompt the development of new therapeutics specifically designed to exploit these differences. In this context, the amino acid depriving enzymes have shown very promising results and proven to be active and very specific against various types of cancers. These therapies involve the depletion of specific amino acids in the blood stream that cannot be synthesised by tumour cells. This happens because these cells have often a defecting enzymatic armamentarium and therefore rely on external supply for those amino acids. Decreasing the concentration of certain amino acids in blood has thus been shown to impair the development or even destroy tumour cells. Normal cells remain unaltered since they are less demanding and/or can synthesize these compounds in sufficient amounts by other mechanisms.

In this chapter, the structure, function, catalytic mechanism and therapeutic application of some amino acid depriving enzymes will be reviewed. Particular attention will be given to enzymes that have potential or are currently used in the treatment of several types of cancer, namely : i) L-asparaginase used for the treatment of acute lymphoblastic leukaemia, ii) L-arginase and L-arginine deiminase that are used in the therapy of hepatocellular carcinomas and melanomas, two diseases that account annually with approximately 1 million of new cases and for which there is currently no efficacious treatment and iii) L-Methioninase with potential to be used in the treatment of breast, colon, lung and renal cancers.

# 1. Introduction

Cancer has become the leading cause of death in the developed world and has remained one of the most difficult diseases to treat. During the last 5 decades, cancer therapy has relied mainly on low molecular weight molecules. However, this class of compounds, very potent in their action, unfortunately lacks tumour specificity and has often unacceptable toxicities toward normal cells. All of these factors are often followed by side effects, which often undermine the effectiveness of the therapeutic treatment in patients (Cantor et al., 2012b). For this reason, the demand for new and improved therapeutics against cancer remains high and at the same time a deeper knowledge regarding the development of tumours is required.

The development of tumour cells is still not fully understood, but many advances have been made in this area. Alterations to cellular metabolism seem to constitute a nearly universal feature in many types of cancers. For instance, many tumours exhibit deficiencies in their enzymatic armamentarium and cannot biosynthesize one or more amino acids that are essential for its development, survival and spread. In order to overcome these problems, tumours rely on the extracellular pool of these amino acids to satisfy protein biosynthesis demands and continue to grow without being affected. This means that if the concentration of those amino acids is decreased, the development of tumour cells can be impaired or even annihilated. At the same time, normal cells remain unchanged since they are less demanding or can synthesize these compounds in sufficient amounts by alternative pathways.

These observations are the basis of the amino acid deprivation cancer therapy that involves the systemic depletion of tumour-essential amino acid. This knowledge is not new and was initially suggested over 50 years ago, but only now it is gaining popularity with the advances in protein engineering technology and, on more sophisticated approaches that enable the study of genetic and metabolic differences between tumour and normal cells(Cantor et al., 2012a).

In these tumours, diet is not enough to achieve a therapeutically relevant level for amino acid depletion and therefore enzymes are used instead. Regrettably, the human genome does not encode enzymes with the pharmacological or catalytic requisites that are essential for the pretended therapeutic purpose. For this reason heterologous enzymes (from other organisms), recombinant and/or engineered human enzymes are used both in animal studies and in clinical trials. Currently the enzymes of bacterial origin are the most frequently used.

The main function of heterologous or engineered human enzymes used in amino acid depletion therapies is to decrease the concentration of a certain amino acid in the blood stream (e.g. converting it into another molecule) that is essential for the development of a tumour. Some examples include a large fraction of hepatocellular carcinomas (HCC) and metastatic melanomas that become apoptotic under conditions where the nonessential amino acid L-arginine (L-ARG) in serum is depleted (Ni et al., 2008), central nervous system cancers that respond to L-methionine (L-MET) deprivation (Tan et al., 2010), and acute lymphoblastic leukaemia (ALL) for which depletion of L-asparagine (L-ASN) has been shown to significantly reduce cancer growth (Rytting, 2010b).

In this review, the structure, function, and catalytic mechanism of therapeutic heterologous enzymes involved in L-ARG, L-MET and L-ASP degradation will be revised. Particular attention will be given to the heterologous enzymes L-arginase, L-arginine deiminase, L-methioninase and L-asparaginase taking into account their current therapeutic potential.

## 2. Amino Acid deprivation enzymes

At least 20 amino acids have been identified in cells and from these the human body cannot manufacture de novo (from scratch) nine which, consequently need to be obtained from the diet. These amino acids are normally denominated as essential and include L-phenylalanine, L-valine, L-threonine, L-tryptophan, L-methionine, L-leucine, L-isoleucine, L-lysine, and L-histidine. Six other amino acids are termed conditionally essential. This means that their synthesis can be carried out by humans but their synthesis can be limited by a variety of factors, such as prematurity in the infant or individuals in severe catabolic distress. These amino acids are L-arginine, L-cysteine, L-glycine, L-glutamine, L-proline and L-tyrosine. The remaining five amino acids are dispensable in humans (also denoted as non-essential), meaning they can be synthesized in the body, e.g., L-aspartic acid, L-asparagine, L-glutamic acid, L-alanine and L-serine.

The optimal conditions for amino acid depletion therapies occur when the tumours are sensitive to one of the non-essential or conditionally essential amino acids. This will ensure that only tumour cells will be affected by the therapy whereas the normal cells remain unaltered since they can synthesize these compounds in sufficient amounts by their own means. The most notable example of target enzymatic amino acid depletion

in cancer therapy is illustrated by the remarkable success of L-asparaginase, in the treatment of childhood ALL, and to a lesser extent, non-Hodgkins lymphoma. Another examples involve the use of arginine deiminase in the HCC, melanoma, and other urea cycle-deficient cancer cells that have a high demand for L-arginine. The essential amino acids can also be used in amino acid depletion therapies, but it is not very common. This only happens in tumours where the normal metabolism of one essential amino acid is disrupted or when there is some sort of defect in the ability to use a certain amino acid. This is for example the case of L-methionine that has been shown to be detrimental for the survival of a variety of tumour tissues, including colon, breast, prostate, ovary, lung, brain, kidney, stomach and bladder cancers as well as larynx melanoma, sarcoma, leukaemia and lymphomas. In this case the action of methionine-γ-lyase has been shown to be successfully used in the treatment of these types of cancers.

Taking into account the therapeutic potential of these enzymes in cancer therapies, they have been deeply studied in the last two decades. These therapies have been showing promising results, but also some problems in clinical trials. In the following sections a detailed analysis of each of these enzymes is done based on the amino acid that they metabolize.

## 2.1. L-asparagine

L-asparagine (ASN) is a non-essential amino acid that is involved in the metabolic control of cell functions in nerve and brain tissue. This amino acid is important in the synthesis of a large number of proteins and plays an important role in the biosynthesis of glycoproteins where it provides key sites for *N*-linked glycosylation. ASN also plays an important role in the metabolism of ammonia, which is toxic in the human body.
Since L-asparagine is a non-essential amino acid, it can be easily synthesized by the cells. The precursor in the biosynthesis of ASN is oxaloacetate which reacts with L-glutamate to form L-aspartate and α-ketoglutarate, in the presence of a transaminase enzyme. The L-aspartate then reacts with L-glutamine to form L-asparagine in an ATP-dependent reaction that is catalysed by L-asparagine synthetase (Richards and Kilberg, 2006).

Certain types of tumours, such as leukemic cells, cannot synthesize ASN because they lack or have deficient expression of the enzyme L-Asparaginase synthase. This means that the proliferation and survival of these tumour cells are dependent on the external supplies of this amino acid. As normal cells can synthesise ASN, this means

that L-asparagine deprivation therapies would mainly interfere with tumour cells that are ASN deficient.

L-Asparaginase (ASNase) is a bacterial enzyme that has been used with great success in leukaemia treatment regimens for four decades. This enzyme performs the opposite reaction of L-asparagine synthetase by hydrolysing ASN to L-aspartic acid and ammonia. By reducing the levels of plasmatic L-asparagine, ASNase starves leukemic cells and promotes tumour cell apoptosis.

The therapeutic role of ASNase was first note by Kidd in 1953,(Kidd, 1953) that found that guinea pig serum had antitumor activity against two kinds of lymphoma in a murine model and lymphosarcoma in rats. This fact was, latter on, attributed to the enzyme ASNase that was present in the serum (Broome, 1961, Broome, 1963). Later on, in 1966, DeLowrey and co-workers using a purified guinea pig serum treated a boy with acute lymphoblastic leukaemia, and obtained an objective response (Dolowy et al., 1966). This turned ASNase into a promissing anticancer drug, and therefore large amounts of the enzyme were required to start new studies and large-scale clinical trials. However this was not an easy task at that time since the isolation and purification of the enzyme was difficult and not all the ASNases possess anti-tumour activity.

L-Asparaginases (EC 3.5.1.1) is a relatively wide spread enzyme, found in many microorganisms such as *Aerobacter*, *Bacillus*, *Pseudomonas*, *Serratia*, *Xanthomonas*, *Photobacterium* (Peterson and Ciegler, 1969), *Streptomyces* (Dejong, 1972), *Proteus* (Tosa et al., 1971), *Vibrio* (Kafkewitz and Goodman, 1974) and *Aspergillus* (Sarquis et al., 2004). After intense research ASNases derived from *Escherichia coli* (Ecoli-ASNase) and *Erwinia chrysanthemi* (Echry-ASNase) turned out to have the best anti-cancer capabilities. Ecoli-ASNase has thereby become the main source of ASNase because it is easy to produce in large amounts. Echry-ASNase is typically reserved for cases of Ecoli-ASNase hypersensitivity.

The *E. coli* expresses two ASNases denoted as Ecoli-ASNase-I and Ecoli-ASNase-II that are found in different intra or extracellular localizations. Ecoli-ASNase-I (cytosolic) has a lower affinity for ASN, while Ecoli-ASNase-II (periplasmic) has a high ASN affinity. Due to the higher affinity of Ecoli-ASNase-II towards L-asparagine, this enzyme is the mostly studied enzyme and the one that is used for oncologic therapeutic purposes.

Ecoli-ASNase-II is a tetrameric protein that is composed by four identical subunits, each one containing 326 amino acid residues. The binding pocket of this enzyme involves residues from both subunits of each dimer, namely, THR12, TYR25, SER58, GLN59, THR89, ASP90 and LYS162 from one subunit (Figure 1 - red chain),

and ASN248 and GLU283 from the other subunit (Figure 1 - grey chain). All of these residues are interconnected by a strong hydrogen bond network, including a water molecule that is structurally conserved. The position of the amino acid residues in the active site suggests several possible pathways for the catalytic reaction, but the almost symmetric location of two threonine residues, THR12 and THR89, observed in the X-ray structure 3ECA above and below carbon C2 of the substrate, suggests that one of them must be directly involved in the reaction (Figure 1) (Swain et al., 1993).

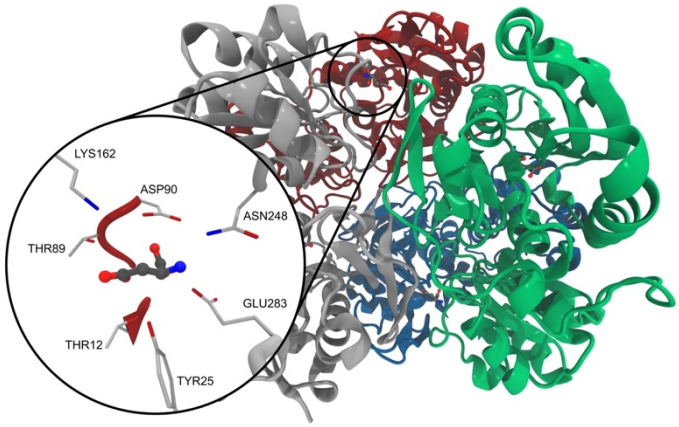



Figure 1 | New Cartoon representation of the co-crystalized structure of ASNase with L-Asparagine (PDB Code: 3ECA). Each subunit is represented with a different colour (blue, grey, red and green) and the substrate in ball and sticks and coloured by CPK. In the close-up of the active site the residues of the active site that are important for the catalytic process as well as the substrate are represented in licorice.

The first step of the mechanism involves the formation of a tetrahedral intermediate (Figure 2a) that results from the nucleophilic attack of the water molecule to α-carbon of the substrate. In the course of this reaction, LYS162 receives the proton from the water molecule and becomes positively charged. The second step (Step 2) of the reaction involves a concerted double proton transfer from LYS162 to THR89 and from THR89 to the substrate, from which results the formation of ammonia and aspartic acid. However, the ammonia molecule is still weakly bound to the substrate at this stage and its dissociation requires an additional step (Step 3). At the end of this process (Step 4), the enzyme is ready for the next turnover (Gesto et al., 2013).

Ecoli-ASNase-II and Echry-ASNase have been established as a general treatment for acute lymphoblastic leukaemia for nearly 40 years. It is used for remission induction and intensification treatment in all paediatric regimens and in the majority of adult treatment protocols. Currently, it is estimated that these enzymes have contributed to significant improvements of therapy outcomes and to achieve complete remission of leukaemia in about 90% of patients.

Notwithstanding its high therapeutic efficacy, the administration of Ecoli-ASNase-II for long periods produces acute allergic reactions and silent immunity, a condition that is characterised by circulating antibodies and rapid clearance of the enzyme from the blood (Henriksen et al., 2015, Avramis et al., 2002, Woo et al., 1998). In order to overcome this problem, a modification of Ecoli-ASNase-II with 5-kDa units of monomethoxypolyethylene glycol was introduced (PEG-Ecoli-ASNase-II) (Rytting, 2010a). The modified enzyme preserves enzyme activity, but decreases immunogenicity of the protein and extends elimination half-life (5 times longer then native enzyme) (Keating et al., 1993).

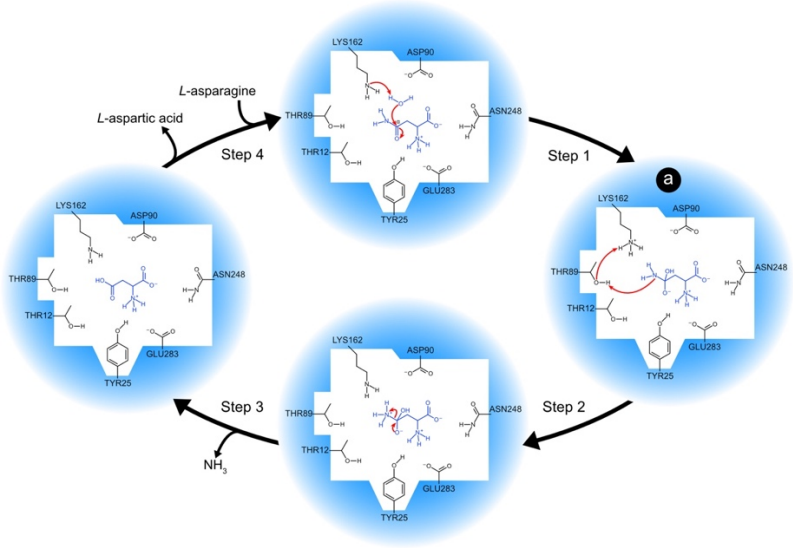



Figure 2 | Schematic representation of the ASNase catalytic mechanism. Each blue circle represents a different reaction intermediate. The black arrows indicate the reaction steps and the red ones the reactions that occur in each step of the mechanism. The important reaction intermediates are labelled with letters (a).

Currently, three L-Asparaginase agents are available. Either these are derived from *E. coli* in its native form (Ecoli-ASNase-II) or as a pegylated enzyme (PEG-Ecoli-ASNase-II) or extracted from *Erwinia chrysanthemi* (Echry-ASNase). Due to the unique anticancer mechanism of action, PEG-Ecoli-ASNase-II and Echry-ASNase have already been approved by US Food and Drug Administration (FDA) for the treatment of several cancers such as acute lymphoblastic leukaemia (Hill et al., 1967), lymphosarcoma (Story et al., 1993) and a few subtypes of non-Hodgkin's lymphoma (Kobrinsky et al., 2001). Ovarian carcinomas and other solid tumours have also been proposed as additional targets for ASNase, with a potential role for its glutaminase activity (Covini et al., 2012).

## 2.2. L-arginine

L-arginine (ARG) is a precursor/mediator of a series of biological pathways, some of them involved in important cellular functions such as in nitrogen metabolism (Rogers and Visek, 1985), creatine, agmatine and polyamine synthesis (Klein and Morris, 1978). In addition, ARG is the major substrate for the production of nitric oxide (Billiar, 1995), stimulates anabolic hormone release, and it is immunostimulatory and thymotrophic (Barbul et al., 1977, Barbul et al., 1983).

The enzyme Argininosuccinate synthetase (ASS) catalyses the condensation of L-Citrulline and L-Aspartic acid to Argininosuccinate which is subsequently converted to ARG and Fumaric acid by Argininosuccinate lyase (ASL) (Figure 3-A) (Haines et al., 2011).

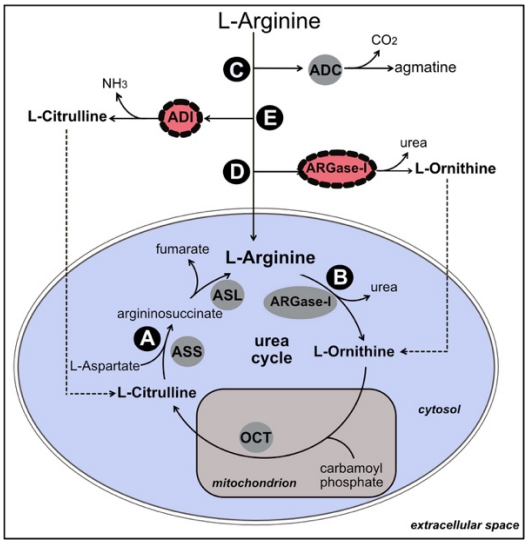



Figure 3 | Schematic representation of the enzymes involved in ARG metabolism and degradation. The enzymes available on the cell have grey background. Recombinant enzymes or from bacterial sources are represented with red background.

In healthy adults the level of endogenous synthesis is sufficient to make ARG a non-essential dietary amino acid (Castillo et al., 1993). However under catabolic stress (e.g. inflammation, infection, etc.), the levels of endogenous synthesis may not be sufficient to meet metabolic demands and ARG becomes an essential amino acid. For this reason, this amino acid is often denoted as a semi-essential amino acid instead of non-essential (Reeds, 2000).

Certain tumour cells such as metastatic melanoma, prostate carcinomas, HCC (Dillon et al., 2004), non-Hodgkin's lymphoma, Hodgkin's lymphoma, pancreatic carcinoma, osteosarcoma and malignant pleural mesothelioma and some breast tumours (Qiu et al., 2015) have an elevated requirement for ARG due to the deficient expression of ASS. This causes ARG autotrophy and therefore these tumours rely only

on exogenous sources of ARG for growth and proliferation. Deprivation of this amino acid is, for this reason, investigated as a novel strategy for cancer therapy and has been showing promising efficacy against the ARG-auxotrophic tumours mention above (Feun et al., 2015, Shen et al., 2003, Wheatley and Campbell, 2003).

ARG can be degraded by several mechanisms, either using enzymes available on the cell, such as L-Arginase (ARGase) (Figure 3-B) or L-arginine decarboxylase (ADC) (Figure 3-C), using human recombinant ARGase-I (Figure 3-D), or using parasite-enzymes from a bacterial source such as Arginine deiminase (ADI, Figure 3-E). From these enzymes, ARGase-I and ADI are the most promising ones since they have therapeutic potential. L-arginine decarboxylase is relatively toxic to normal cells and therefore it is not used for therapeutic purposes (Savaraj et al., 2010).

### 2.2.1. L-Arginase

ARG can be degraded directly to L-Ornithine by ARGase (EC 3.5.3.1), an enzyme of the urea cycle (Figure 3-B). Thus the insertion of ARGase on the extracellular moiety could be used to decrease the levels of ARG in cells. However, under salvage conditions, higher concentration of L-Ornithine (product of the reaction of ARGase) can be converted back to L-Citrulline by Ornithine transcarbamyl transferase and then recycle back to ARG by ASS/ASL (Figure 3-A). This feedback mechanism makes normal cells and many tumours resistant to ARGase inhibition. Fortunately, there are many tumours that present deficient expressions of OCT and/or ASS enzymes and in these cases ARGase can be efficiently used to decrease ARG intracellular levels and therefore tumour growth (Cheng et al., 2007).

ARGase is found in organisms from the five kingdoms (Reczkowski and Ash, 1992). In most mammals, two isoforms of this enzyme exist and are denoted as types I and II. They share approximately 60% amino acid sequence homology and differ in their expression regulation, tissue distribution, subcellular locations, immunologic reactivity, physiologic function and certain enzymatic properties (Vockley et al., 1996, Spector et al., 1983, Jenkinson et al., 1996, Herzfeld and Raper, 1976, Dizikes et al., 1986). ARGase-I is a cytosolic enzyme located predominantly in the liver (Krebs, 1973). ARGase-II is a mitochondrial enzyme that does not appear to function in the urea cycle and is more widely distributed in numerous tissues (Morris et al., 1997).

In mammals, ARGase-I is present in the liver and it is one of the most important enzymes involved in nitrogen metabolism. It is involved in the principal route for disposal of excess nitrogen resulting from amino acid and nucleotide metabolism. The flux of nitrogen

through this pathway is considerable, given that the average individual excretes about 10 kg of urea per year (Ash et al., 1998).

In humans, h-ARGase-I is an oligomer of 107 kDa and shows optimum pH at 9.3 and has a Km for ARG about 10.5 mM (Beruter et al., 1978) . The key structural features of mammalian ARGase-I include a binuclear Manganese cluster (Mn), and an S-shaped tail composed by 19 amino acids at the C terminus of each monomer. The S-shaped tails are located at the subunit-subunit interface, and 54% of the inter-subunit interactions are mediated by this region of the protein (Lavulo et al., 2001).

The catalytic mechanism of this enzyme is still not fully understood but several suggestions have already been proposed, some of them more consensually accepted (Christianson and Cox, 1999, Cox et al., 2001, Leopoldini et al., 2009) than others (Khangulov et al., 1998). Among the active site residues, GLU277 is apparently implicated in the substrate binding through the establishment of electrostatic interactions with the guanidium group of the substrate. This interaction appears to orientate the subtract in a favourable position with respect to the metal-bound hydroxide anion with whom it must react with (Cavalli et al., 1994, Kanyo et al., 1996, Scolnick et al., 1997). Both Manganese ions are coordinated to a hydroxide anion and with two Aspartates (ASP232 and ASP128) and one Glutamate (GLU277). These structural features can be observed in the crystallographic structure with the PDB code 3GMZ, where the human ARGase is in complex with L-ornithine. (Figure 4) (Ilies et al., 2011).

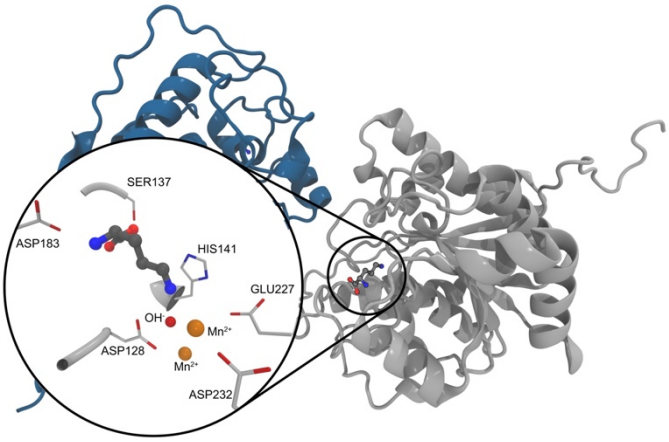



Figure 4 | New Cartoon representation of the co-crystalized structure of h-ARGase with L-Ornithine (PDB Code: 3GMZ). Each subunit is represented with a different colour (blue and grey) and the substrate in ball and sticks and coloured by CPK. In the close-up of the active site are represented: the residues of the active site that are important for the catalytic process in licorice and, the oxygen atom from the hydroxide ion and the two magnesium ions in VDW.

According to (Leopoldini et al., 2009) the hydrolysis of ARG mediated by ARGase-I is kinetically favoured when the ASP128 residue is deprotonated. The hydrolysis starts with the binding of ARG to the Mn(II) ions through the gaunidium groups.

These interactions place the substrate in a good position to undergo nucleophilic attack by the bridging hydroxy anion to carbon Cζ (Step 1) from which results a tetrahedral intermediate (Figure 5a). In the second step of the mechanism, ASP128 acts as an acid/base catalyst, abstracting a proton from the hydroxyl group and shuttling it to the nitrogen Nε of the substrate (Step 2). This step is then followed by the bond break between nitrogen Nε and carbon Cζ of the substrate from which results L-ornithine and urea (Step 3). In the following step (Step 4), a water molecule enters to bridge the binuclear Mn(II) cluster, causing the urea product to move to a terminal coordination site on Mn(II). HIS141 appears to facilitate product release by serving as a proton shuttle from bulk solvent to nitrogen Nε of ornithine and regenerating the nucleophilic metal-bridging hydroxide anion (Step 5). Once this step is complete ornithine dissociates, the enzymatic turnover takes place and the enzyme is ready for a new catalytic cycle (Step 6).

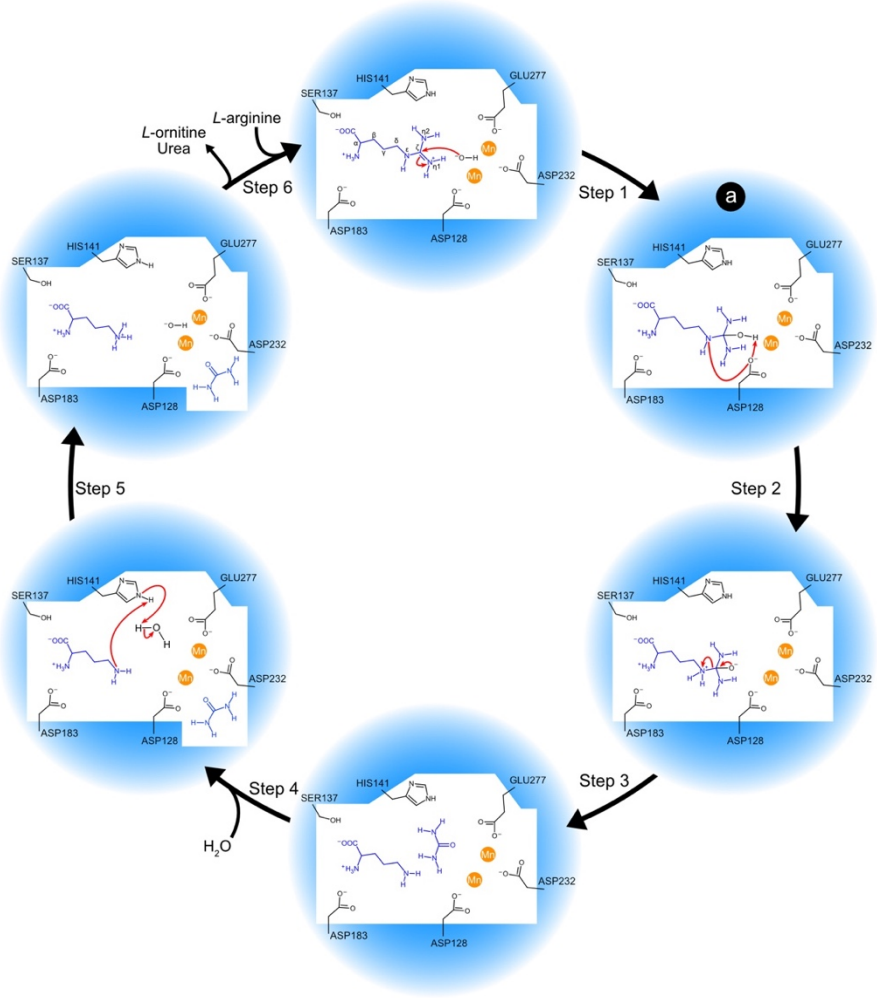



Figure 5 | Schematic representation of the ARGase catalytic mechanism. Each blue circle represents a different reaction intermediate. The black arrows indicate the reaction steps and the red ones the reactions that occur in each step of the mechanism. The important reaction intermediates are labelled with letters (a-c).

h-ARGase-I has been tested with success in experimental animals since 1950, and showed promising activity *in vitro* conditions(Simon-Reuss, 1953, Wiswell, 1951). However, the therapy failed to produce major responses when tested *in vivo* conditions. The reason for this failure was related with the low affinity of this enzyme for ARG with the consequent requirement of a large quantity of enzyme in order to obtain the expected therapeutic results. In addition, this enzyme has a short circulatory half-life (only a few minutes) (Savoca et al., 1979) and its optimal pH is very high (9.5) that cannot be achieved under physiological conditions. All of these drawbacks declined the use of h-ARGase-I as an antitumor agent in the last decades (Currie et al., 1979). In spite of these drawbacks it was shown, in 2010 that a pegylated recombinant human arginase I (PEG-r-h-ARGase-I) impairs the proliferation of malignant T cells in acute lymphoblastic T cell leukaemia *in vivo* and *in vitro* studies (Hernandez et al., 2010).

Later on, a pegylated cobalt-modified enzyme was developed and showed that the substitution of the Mn(II) metal center in h-ARGase-I by Co(II) (Co-r-h-ARGase-I) turns the enzyme more efficient for ARG hydrolysis and lowers the IC50 in a variety of malignant cell lines, including melanomas and hepatocellular carcinomas *in vitro* assays (Stone et al., 2010). The onlysame group also reported a modification of Co-r-h-ARGase-I with PEG-5K esters that shown an increase retention of the enzyme in circulation by about 2 orders of magnitude. The use of PEG(5K)-Co-h-ARGase-I was later on be found to be effective in the treatment of hepatocellular and pancreatic carcinomas in xenograft models (Glazer et al., 2011).

### 2.2.1. L-arginine deiminase.

L-arginine deiminase (ADI, EC 3.5.3.6) is widely distributed among prokaryotic organisms and some anaerobic eukaryotes, but has never been identified in humans or other higher eukaryotes (Figure 3-E). The organisms that employ this pathway use ADI to convert ARG to L-Citrulline and ammonia (Horn, 1933). L-Citrulline is then degraded further, forming ATP (Adenosine Tri-Phosphate), carbon dioxide, and L-ornithine (Arena et al., 1999). The ADI pathway is thus used by these organisms to get energy, carbon, and nitrogen. In addition, it has also been suggested that this pathway can also protect some bacteria from acidic conditions, by the production of ammonia (Casianocolon and Marquis, 1988).

The interest in ADI for therapeutic purposes began in the 90s when Takaku and co-workers successfully used this enzyme to inhibit the growth of several murine and

human tumour cell lines *in vitro* and *in vivo*, by exhausting the supplies of ARG (Takaku et al., 1995, Takaku et al., 1992). From there one, several studies confirmed the potential of ADI, purified from Mycoplasma, as an anti-tumour agent in different types of tumours (Takaku et al., 1992) (Sugimura et al., 1992) (Ashikaga et al., 1994). Later on, it was also found that beyond the anti-tumour activity, ADI has an anti-angiogenic activity via suppression of nitric oxide generation (Beloussow et al., 2002, Park et al., 2003, Yoon et al., 2007). The cooperative anti-proliferative and anti-angiogenic activities of ADI are believe to be the key assets that turn ADI very effective in the treatment of several tumours and other diseases.

ADI has been structurally and pharmacologically characterized from various bacteria, but the most studied enzyme is ADI from *Mycoplasma arginini* take into account its therapeutic potential (Zhang et al., 2015). ADI from *M. arginini* has 410 amino acids and is composed of two identical subunits with a molecular weight of 45 kDa each and has a isoelectric point of 4.7. Its Vmax value and Km value for ARG is estimated to be 50 units/mg protein and 0.2 mM, respectively. It exerted maximal enzyme activity at pH 6.0-7.5 and at 50 °C (Takaku et al., 1992).

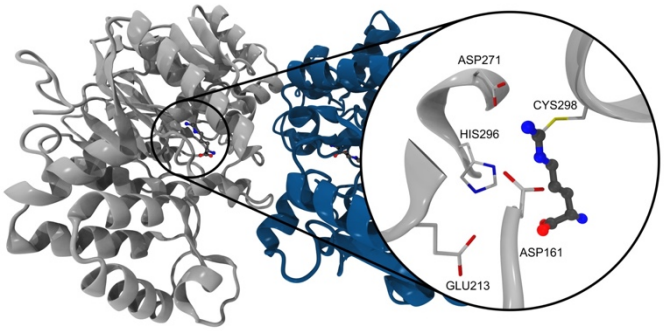



Figure 6 | New Cartoon representation of the co-crystalized structure of ADI homodimer structure with a tetrahedral reaction intermediate (PDB Code: 1S9R). Each subunit was represented with a different colour (blue and grey) and the reaction intermediate in ball and sticks and coloured by CPK. In the close-up of the active site the residues of the active site that are important for the catalytic process are represented in licorice as well as the reaction intermediate that should be formed after the first step of the mechanism.

A proposal for the catalytic mechanism of ADI from M. arginini was published by Das and co-workers in 2004 (Das et al., 2004). The crystallographic structures available on the protein databank show that the active site is deeply buried on the protein surface and the substrate binds with the guanidium group pointing towards two aspartates (ASP161 and ASP271). These interactions endorse a specific orientation of the substrate that becomes aligned with HIS296 and CYS398, two residues that play an

important role in the catalytic process . These structural features are represented in Figure 6 (PDB code 1S9R) (Das et al., 2004). The importance of the CYS398 in the catalytic reaction is well established since it has already been demonstrated trough structural studies, mutagenesis and thiol modifying agents (Das et al., 2004) (Smith et al., 1978, Lu et al., 2004) (Lu et al., 2006, Galkin et al., 2005).

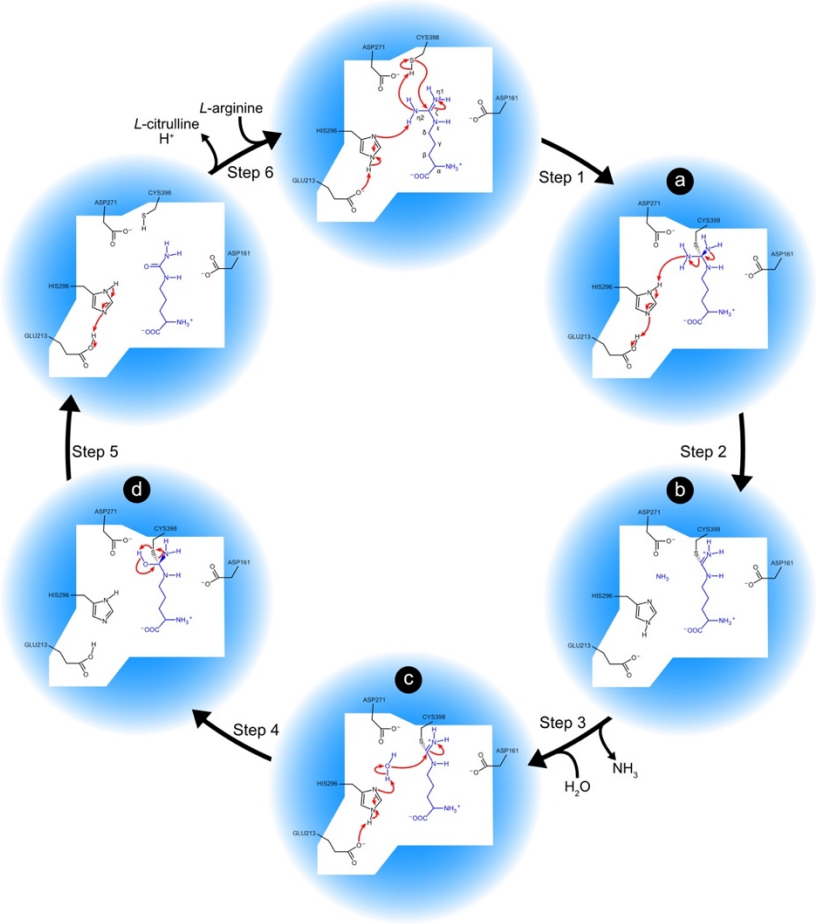



Figure 7 | Schematic representation of the ADI catalysis mechanism. Each blue circle represents a different reaction intermediate. The black arrows indicate the reaction steps and the red ones the reactions that occur in each step of the mechanism. The important reaction intermediates are labelled with letters (a-d).

In the first step of the mechanism there is a nucleophilic attack of CYS398 to carbon Cζ of the substrate. Simultaneously the proton from CYS398 migrates to nitrogen Nη2 of the substrate. At the same time, one proton from the same nitrogen migrates to HIS296 and the proton from HIS296 is transferred to GLU213. In the end of the first step, a tetrahedral intermediate (Figure 7a) is generated with carbon Cζ of the substrate covalent bonded to CYS398.

In the second step of the catalytic mechanism occurs the formation of ammonia through the migration of one proton from HIS296 to Nη2 of the substrate. Simultaneously, one proton is transferred from GLU213 to HIS296. In the end of this reaction, the reaction

intermediate (Figure 7b) remains covalently bonded to CYS398 and one molecule of ammonia is released from the active site.

The next steps of the mechanism (Figure 7: step 4-6) involve the release of the product of the reaction (L-Citrulline) and the enzymatic turnover. This requires three sequential steps and involves the reaction of one water molecule that is in close proximity to ASP271 and HIS296. In the forth step of the catalytic mechanism occurs the nucleophilic attach of the water molecule to carbon Cζ of the reaction intermediate (Figure 7c). At the same time one proton from the water molecule is transferred to HIS296 and one proton migrates from HIS296 to GLU213. The reaction is endorsed by the close proximity of ASP161 that establishes two hydrogen bonds with the NH groups of the intermediate that turns carbon Cζ more electrophilic. In the end of this reaction, another tetrahedral intermediate is generated (Figure 7d) still bonded to CYS398 but now with a hydroxyl group attached to carbon Cζ instead of an amino group that was found after step 1. Step 5 involves the cleavage of the covalent bond between CYS398 and the substrate. This requires the proton transfer from hydroxyl group of the reaction intermediate to CYS398. Simultaneously, occurs the proton transfer from GLU213 to HIS296. In the end of this reaction, L-Citrulline is generated and it is ready to leave the active site. The final step of the reaction (Step 6) involves the enzymatic turnover that requires HIS398 to lose a proton to the solvent or any neighbor active site residue.

In recent years there has been an increased interest in ADI as a potential antineoplastic therapy since various studies have demonstrated disturbances in both ARG synthesis and metabolism in malignant processes. To date several types of cancer have already been shown to be sensitive to arginine deprivation via ADI, such as melanoma, lung cancer, renal cell carcinomas, and HCC (Liu et al., 2014). These tumours do not express ASS and are therefore highly sensitive to ARG deprivation and optimal targets for ADI chemotherapies. In spite of the promising results, the application of these treatments presents some limitations. The therapeutic efficiency of ADI is limited to tumours that do not express ASS and/or have an inactivated L-Citrulline to ARG recycling pathway (Ensor et al., 2002, Qiu et al., 2014). In addition, ADI has a short serum half-life and when used in prolonged treatments is highly immunogenic. In order to overcome some of these limitations a pegylated form of ADI was formulated, PEG(20)-ADI (Molecular weight: 20 kDa), that serves to reduce the immunogenicity of the enzyme while greatly improving its pharmacokinetic half-life in serum (Ni et al., 2008, Feun and Savaraj, 2006, Feun et al., 2008). This enzyme revealed a similar efficiency *in vitro* assays with melanomas and HCC, comparing with native ADI, however, it shows a much more effectiveness i*n vivo* assays (Ensor et al., 2002). Currently, the anti-tumour activity

of PEG(20)-ADI has been also observed in pancreatic (Bowles et al., 2008), prostate (Kim et al., 2009), small cell lung (Kelly et al., 2012), head, neck (Huang et al., 2012) and breast cancers (Qiu et al., 2014), lymphoma (Delage et al., 2012, Huang et al., 2012), myxofibrosarcomas (Huang et al., 2013), melanoma (Feun et al., 2012) and glioblastoma (Syed et al., 2013, Qiu et al., 2014).

The therapeutic efficacy of PEG(20)-ADI has been validated through phase I/II clinical trials in advanced HCC (Glazer et al., 2010, Izzo et al., 2004, Yang et al., 2010) and melanoma patients (Ascierto et al., 2005, Feun et al., 2012, Ott et al., 2013). Additionally, there are several ongoing phase I and/or II clinical trials using PEG20-ADI in patients with advanced non-Hodgkin's lymphoma, acute myeloid leukaemia, malignant pleural mesothelioma, prostate and non-small cell lung, advanced gastrointestinal, HER2 (Human Epidermal growth factor Receptor 2) negative metastatic breast, advanced pancreatic, and small cell lung cancers.

PEG(20)-ADI has also been reported as a potentially better therapeutic agent for the treatment of leukaemia than ASNase, which has been used for the treatment of acute leukaemia for over 20 years. ADI can inhibit the growth of cultured leukaemia cells at concentrations of 5–10 ng/ml, which are about 20–100 times lower than those of ASNase (Feun et al., 2008). In addition, *in vivo* studies showed that ADI treatment does not have serious side-effects such as anaphylactic shock, coagulopathies, and liver toxicity, which have been reported for ASNase (Vrooman et al., 2010).

## 2.3. L-methionine

L-methionine (MET) is a sulphur containing amino acid that plays many key roles in cells (Guedes et al., 2011) (Nozaki et al., 2005). In every cell, MET is essential for protein synthesis and it is the precursor of many compounds, such as Glutathione (a tripeptide that reduces reactive oxygen species, thereby protecting cells from oxidative stress), polyamines spermine and spermidine (which are detrimental on nuclear and cell division) and S-adenosylmethionine (the major source of methyl groups in a myriad of biological and biochemical events) (Ho et al., 2011) (Cellarier et al., 2003a) (Thivat et al., 2007, Davis and Uthus, 2004). One of the most relevant biological role of MET in the cell is DNA methylation, where it serves as precursor for S-adenosylmethionine synthesis. DNA methylation is a common epigenetic signalling tool that cells use to control gene expression and has been associated with a number of key processes including genomic imprinting, X-chromosome inactivation, repression of repetitive elements, and carcinogenesis (Geiman and Muegge, 2010) (Ramani et al., 2010).

In spite of its importance in cells, MET is a special essential amino acid as its long-term dietary deprivation is still compatible with life. This happens because unlike other essential amino acids that must be ingested, MET can be recycled by re-methylation of homocysteine (*de novo* synthesis), either by L-methionine synthase (MetS) or in the liver by betaine-homocysteine methyltransferase (BHMT) (Cellarier et al., 2003a, Hoffman, 1984, Guo et al., 1993a). In addition, MET can also be obtained by a *salvage* pathway in which MET is generated from 5′-methylthioadenosine towards Methylthioadenosine phosphorylase (MTAP). This is true for all normally functioning cells, however, numerous malignant cell lines, such as breast, lung, colon, kidney and bladder cancers, melanoma and glioblastoma (Breillout et al., 1990, Hoffman, 1985, Kokkinakis et al., 2001, Poirson-Bichat et al., 2000, Cellarier et al., 2003b) do not have a functional methionine cycle intact. As a result, they cannot synthesize MET and in the absence of an exogenous supply they experience growth inhibition or eventually they die.

The mechanisms responsible for MET dependence in malignant cell lines are still not fully understood. Several studies have shown that MET dependent tumour cell lines presented relatively low amounts of L-methionine synthase or/and have deficient methionine synthase activity (Kenyon et al., 2002, Christa et al., 1986, Fiskerstrand et al., 1994). Other studies emphasized the importance of the inhibition of the salvage MET pathway in which MTAP is a key enzyme and is also responsible for the production of polyamines which are critical for cell proliferation. Indeed, the loss of MTAP activity through gene deletion is a common feature in many kinds of human cancers including non-small cell lung cancer, glioma, T-cell acute leukaemia, bladder cancer, osteosarcoma and endometrial cancer (Nobori et al., 1993, Traweek et al., 1988, Subhi et al., 2003, Albers, 2009). In contrast, other studies suggest that MET dependence would result not from an enzymatic deficiency but rather on an increased requirement for this amino acid. In fact, MET requirements of tumour cells are higher than that of normal cells because of increased protein synthesis and DNA trans-methylation reactions. This promotes multiple biochemical reactions necessary for fast growth of tumour cells, and can change DNA expression. Indeed, in several tumours, a variety of growth inhibitory and pro-apoptotic genes are transcriptionally silenced as a result of DNA hyper-methylation (Zingg and Jones, 1997) (Baylin et al., 1998).

The observation that some human tumours are MET dependent *in vivo* presents a therapeutic target for cancer growth control and has been studied in great detail in the last two decades (Hoshiya et al., 1995). As MET is sourced mainly from diet, initial studies observed that simple methionine dietary restriction caused regression of a variety

of animal tumours and inhibits metastasis in animal models (Breillout et al., 1987, Breillout et al., 1990, Guo et al., 1993b, Goseki et al., 1987, Millis et al., 1998). Phase I clinical trials of dietary methionine restriction for adults with advanced cancer indicate that dietary methionine restriction is safe and provided promising results. However, reduction of methionine levels in humans solely by dietary intervention presents several limitations as MET is practically ubiquitous for all food and the price of such diet is about US$ 1000 per month.

A more attractive pharmacological approach to lowering MET levels *in vitro* and *in vivo* involves the use of MET degrading enzymes. L-methionine-α-amino-γ-mercaptoethane lyase or methioninase (MGL, EC 4.4.1.11) is an enzyme that specifically degrades L-methionine to α-ketobutyrate, methanethiol and ammonia (Kreis and Hession, 1973) very efficiently. In addition, it can also degrade L-homocysteine, a key factor that prevents endogenous MET production(Sato and Nozaki, 2009).
MGL is found in bacterial species and in fungi, but it is absent in mammals. (Sharma et al., 2014). This enzyme was originally purified from *Clostridium sporogene*, but later on it was found that MGL from *Pseudomonas putida* (Pp-MGL) yielded a more stable enzyme with a relatively low Km (Esaki and Soda, 1987). Due to problems in purification of Pp-MGL, the MGL gene, isolated from *P. putida*, is now cloned and expressed in great quantities from *Escherichia coli* (r-Pp-MGL) (Hori et al., 1996).

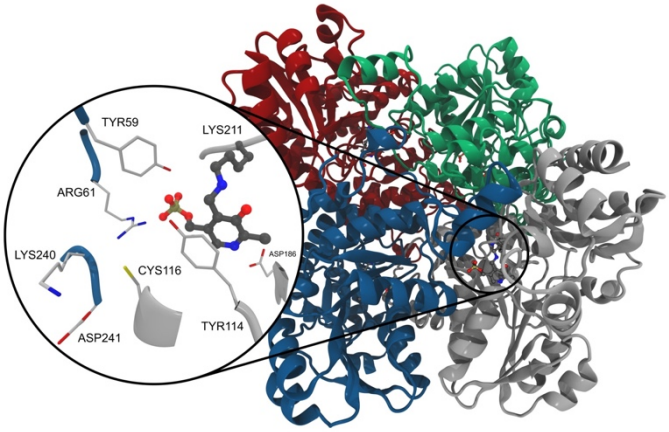



Figure 8 | New Cartoon representation of the co-crystalized structure of the MGL homotetramer structure with the PLP cofactor (PDB Code: 2O7C). Each subunit was represented with a different colour (red, green, blue, grey). In the close-up of the active site it is shown the internal aldimine in ball and sticks, where PLP cofactor covalent bonded to an active site LYS. The other active site residues that are important for the catalytic process are represented in licorice.

Pp-MGL has 389 to 441 amino acids, and it is found as a homotetramer with a molecular weight of about 149 to 173 kDa (Nakayama et al., 1984). The active site of this enzyme is located at the interface of the two neighbouring subunits, each one containing one Pyridoxal 5'-phosphate (PLP) cofactor that is required for the catalytic

process (Toney, 2005, Motoshima et al., 2000, Kudou et al., 2007a) (Mamaeva et al., 2005). The cofactor is stabilized by six residues TYR59, ARG61, TYR114, CYS116, LYS240, and ASP241 that have been found to be important for catalysis (Figure 8) (Kudou et al., 2007a).

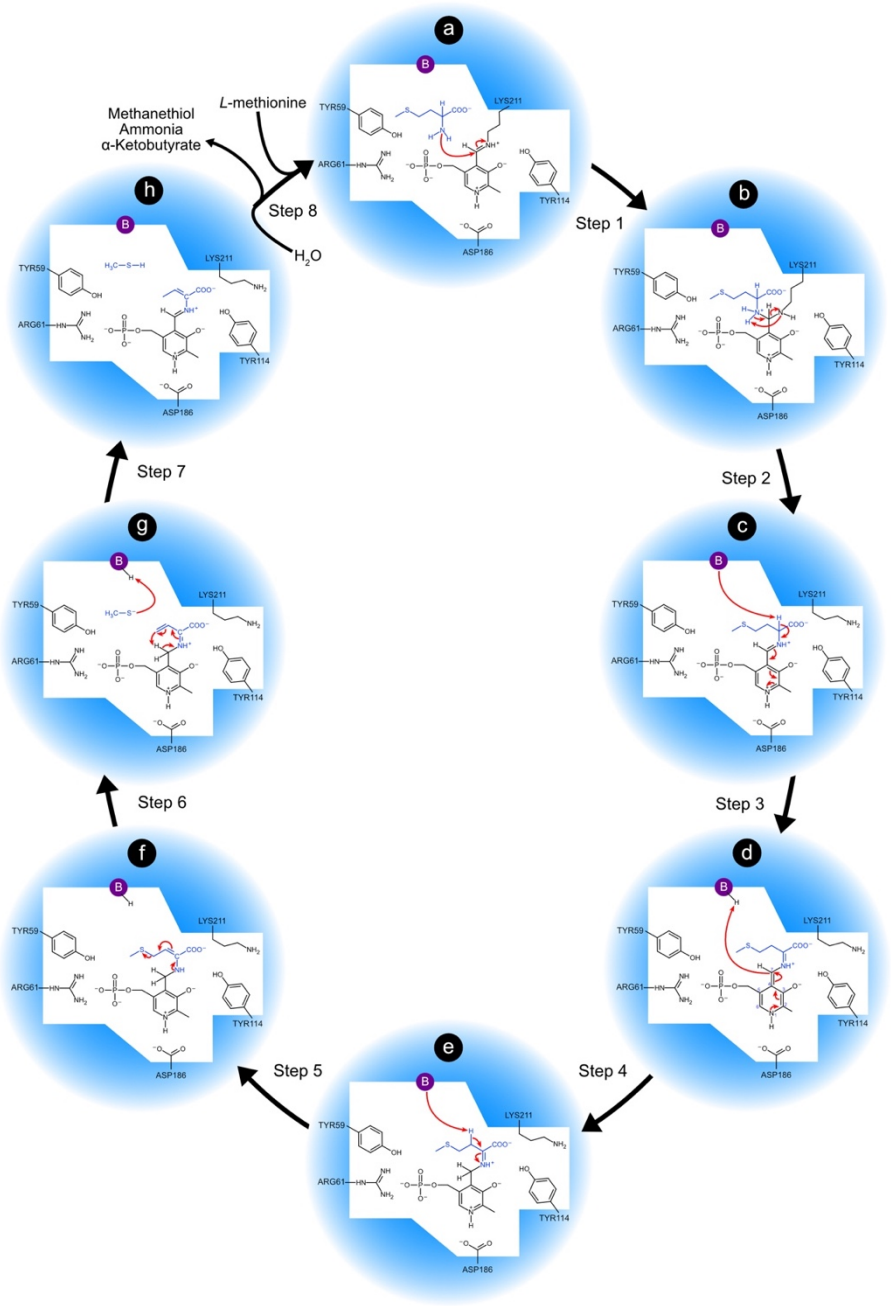



Figure 9 | Schematic representation of the MGL catalysis mechanism. Each blue circle represents a different reaction intermediate. The black arrows indicate the reaction steps and the red ones the reactions that occur in each step of the mechanism. The important reaction intermediates are labelled with letters (a-h).

MGL is classified into the γ-family of PLP-dependent enzymes. However, it possesses a key cysteine (CYS116) in the active site that is not conserved in other PLP

γ-family enzymes. Mutagenic studies revealed that this active site CYS is essential for substrate specificity (Sato et al., 2008) (McKie et al., 1998) (Kudou et al., 2008, Kudou et al., 2007a) and also for enzymatic activity (Inoue et al., 2000, Kudou et al., 2008).

The enzymatic mechanism of MGL is proposed based on the knowledge of other PLP γ-family enzymes and the enzymological analyses of *P. putida* wild-type and mutant forms (Fukumoto et al., 2012, Kudou et al., 2007a) (Cerqueira et al., 2011, Oliveira et al., 2011). Similar to all PLP-dependent enzymes, in the steady state the PLP cofactor is covalently bonded to an active site LYS forming what is generally known as an internal aldimine (Figure 9a). When MET is available in the active site of MGL it binds to PLP (Step 1) generating a geminal diamine (Figure 9b). Afterwards, the cleavage of the covalent bond between the PLP cofactor and the active site LYS211 takes place (Step 2) and the substrate becomes covalent bonded to the PLP cofactor through the primary amino group forming an external aldimine (Figure 9c). The next step (Step 3) involves the abstraction of one proton from carbon Cα of the reaction intermediate by an active site base resulting in the formation of a quinonoide intermediate (Figure 9d). Later on, carbon 7 from PLP is protonated by an active site base (Step 4) from which results a ketamine intermediate (Figure 9e). An active site base then abstracts a second proton from carbon Cβ (Step 5) and an enamine intermediate is generated (Figure 9f) that triggers the γ-elimination (Step 6) and the generation of unprotonanted Methanethiol (Figure 9g). Methanethiol is subsequently protonated by the active site base and an Aminocrotonate intermediate is obtained (Figure 9h) (Step 7). In the final step of the catalytic process (Step 8), a water molecule mediates the α-Ketobutyrate and Ammonia generation and the concomitant release of Methanethiol. At the same time, the enzymatic turnover takes place recovering the steady state of the enzyme, i.e., internal aldimine in which the active site LYS is covalent bonded to the PLP cofactor. (Inoue et al., 2000, Toney, 2005, Sato et al., 2008) (Kuznetsov et al., 2015).

The therapeutic manipulation of r-Pp-MGL to deplete plasma MET has been widely investigated in the last two decades (Jeanblanc et al., 2005) with promissing results in a number of L-methionine-dependent cancer cell lines including lung, colon, kidney, brain and prostate cancers, melanoma and fibrosarcoma (Miki et al., 2000a) (Yamamoto et al., 2003) (Tan et al., 1999, Tan et al., 1997, Miki et al., 2000b, Miki et al., 2001, Tan et al., 1996, Smiraglia et al., 2001). Preclinical studies in animal models have also demonstrated the potential of r-Pp-MGL to efficiently inhibited the growth of Yoshida sarcomas and a human lung tumour in nude mice xenograft models (Tan et al., 1996).
Several studies have also shown promising results combining the chemotherapeutic effect of r-Pp-MGL with other compounds, such as, cisplatin, 5-fluorouracil (5-FU), 1-3-

bis(2-chloroethyl)-1-nitrosourea (BCNU), cisplatin, and vincristine (Tan et al., 1999). These studies have shown synergic effects in mouse models of colon, lung and brain cancers and glioblastoma(Yoshioka et al., 1998, Hu and Cheung, 2009, Kokkinakis et al., 2001, Tan et al., 1999, Hoshiya et al., 1997). Other authors have also studied the administration of r-Pp-MGL with Selenomethionine (a nontoxic compound). These studies have shown that r-Pp-MGL catalyses the conversion of Selenomethionine into Methylselenol, and the later compound is able to oxidate thiol groups and generate toxic superoxide. This compound is very toxic to cells and can cause apoptosis via a mitochondrial pathway (Guo et al., 1993a, Kim et al., 2007). As the in cancer cells the metabolism is faster than in normal cells this approach has the advantage of being target specific.

In spite of the initial success of r-Pp-MGL for cancer therapy it presents some problems that precludes an easy manipulation, namely, the instability of r-Pp-MGL in plasma caused by immunogenicity, proteolysis, oxidation of the active site residues, release of PLP and the instability of the active dimer (Kudou et al., 2007b). Modification of this enzyme with polyethylene glycol (PEG) has reduced immunogenicity, lengthen the half-life of the recombinant enzyme, and possibly inhibited proteolytic enzymes that could degrade MGL (Tan et al., 1998). The intravenous administration of PEG-r-Pp-MGL was tested in phase I clinical trials. Patients in these trials experienced no significant toxicity, and plasma MET levels fell dramatically. However, the authors did not assess the anti-tumour activity of this treatment and to date no other clinical trials have been published (Sun et al., 2003a, Yang et al., 2004b, Yang et al., 2004c).

## 3. Conclusions

A prerequisite for making an effective medication in the treatment of cancer is that some fundamental difference between normal cells and cancer cells must be defined. An optimal chemotherapeutic agent must thus exploit this cellular difference in such a way that normal cells are spared and only cancer cells are injured. In this context, the amino acid depriving enzymes are promising anti- cancer drugs that have proven to be active and very specific against various types of cancers. Their mode of action is simple: they decrease the concentration of certain amino acids in the blood stream and thus impair the development or even destroy tumor cells that are autotrophic for those amino acids. Normal cells remain unaltered since they are less demanding and/or can synthesize these compounds in sufficient amounts by other mechanisms.

ASNase is probably the best-known example of amino acid depletion chemotherapies using enzymes. The application of this enzyme in the treatment of childhood acute lymphoblastic leukaemia has already been approved by FDA and shows very good results. ARG deprivation by ADI is a novel approach to target tumors that lack argininosuccinate synthetase (ASS) expression. ADI showed promising results during initial clinical trials for the therapy of HCC and melanomas and has currently been granted by FDA and EMEA for the drug Orphan Drug status for the potential treatment of these diseases(Shen and Shen, 2006). Some studies also suggest that ADI has potential to become a better therapeutic agent for the treatment of leukaemia than ASNase in the treatment of leukaemia(Gong et al., 2000, Muller and Boos, 1998).
Currently, the major drawback of ADI and ASNase for therapeutic purposes is the heterologous origin. As these enzymes are derived from microorganism, they are immunogenic in humans and have very short circulating half-life. The chemotherapeutic treatments using these enzymes thus require multiple injections to produce the desired effect and from which results a variety of adverse effects including hypersensitivity reactions, anaphylactic shock, and the inactivation and clearance of the enzyme itself (Schellekens, 2002). In order to overcome these problems and allow a safer application in cancer treatments, these enzymes have been linked to PEG to improve the pharmacokinetic and pharmacodynamic properties. The therapeutic results of these modified enzyme are much safer and have prompted the use of ADI in many clinical trials and the approval of ASNase by FDA and EMEA for the treatment of some types of tumours.

MGL is another heterologous enzyme discussed in this review. Similarly to ASNase and ADI, this enzyme presents several immunogenic effects, but contrarily to

those enzymes the pegylated form does not seem to overcome this problem. Several attempts were made to decrease the toxicity of this enzyme by the co-administration of pyridoxal 5-phosphate and oleic acid or dithiothreitol (Cantor et al., 2011) (Sun et al., 2003b). However, several side effects persist and its utilization in cancer therapies continues to be inappropriate(Miki et al., 2001, Yang et al., 2004a).

From the discussed enzymes, ARGase is the only enzyme that is from human origin. This means that it presents low immunogenicity, a great advantage over ADI in the treatment of cancer that are autotrophic for ARG. Indeed, several authors consider ARGase a better chemotherapeutic agent for cancer treatment than ADI because apart from its low immunogenicity, this enzyme also is effective in tumours for which ADI is ineffective (although in these cases OCT cannot be expressed). Its major disadvantage when compared with ADI is its high Km and high optimum pH, which requires large quantities of enzyme in order to accomplish the expected therapeutic results. Actually this problem may be overcome by the utilization of cobalt-substituted ARGase, but more studies are required to acquire its potential in cancer therapy.

Despite the problems arising from the use of amino acid depleting enzymes in amino acid depriving chemotherapies, the recent progresses made in this field will undoubtedly facilitate the pursuit of new treatments to combat cancers. ASNase was the first enzyme approved by FDA and EMEA for the chemotherapeutic treatment of leukaemia with remarkable success. Taking into account the on-going clinical trials made with other enzymes such as, ARGase or ADI, and the promising results that are being obtained, one might expect that the clinical approval of these enzymes should not be too distant in the future and will have undoubtedly a profound influence on human health.

## 4. Acknowledgements

This work has been funded by FEDER/COMPETE and the Fundação para a Ciência e a Tecnologia (FCT) through projects IF/01310/2013 and UID/Multi/04378/2013.

# 5. Bibliography

ALBERS, E. 2009. Metabolic characteristics and importance of the universal methionine salvage pathway recycling methionine from 5'-methylthioadenosine. *IUBMB Life,* 61**,** 1132-42.

ARENA, M. E., SAGUIR, F. M. & DE NADRA, M. C. M. 1999. Arginine, citrulline and ornithine metabolism by lactic acid bacteria from wine. *International Journal of Food Microbiology,* 52**,** 155-161.

ASCIERTO, P. A., SCALA, S., CASTELLO, G., DAPONTE, A., SIMEONE, E., OTTAIANO, A., BENEDUCE, G., DE ROSA, V., IZZO, F., MELUCCI, M. T., ENSOR, C. M., PRESTAYKO, A. W., HOLTSBERG, F. W., BOMALASKI, J. S., CLARK, M. A., SAVARAJ, N., FEUN, L. G. & LOGAN, T. F. 2005. Pegylated arginine deiminase treatment of patients with metastatic melanoma: results from phase I and II studies. *J Clin Oncol,* 23**,** 7660-8.

ASH, D. E., SCOLNICK, L. R., KANYO, Z. F., VOCKLEY, J. G., CEDERBAUM, S. D. & CHRISTIANSON, D. W. 1998. Molecular basis of hyperargininemia: Structure-function consequences of mutations in human liver arginase. *Molecular Genetics and Metabolism,* 64**,** 243-249.

ASHIKAGA, T., WANG, Z., YAMAMOTO, M., YAMASAKI, M., MAGAE, J. & NAGAI, K. 1994. Development and characterization of macrophage hybridomas derived from murine peritoneal exudate cells. *Biosci Biotechnol Biochem,* 58**,** 839-42.

AVRAMIS, V. I., SENCER, S., PERICLOU, A. P., SATHER, H., BOSTROM, B. C., COHEN, L. J., ETTINGER, A. G., ETTINGER, L. J., FRANKLIN, J., GAYNON, P. S., HILDEN, J. M., LANGE, B., MAJLESSIPOUR, F., MATHEW, P., NEEDLE, M., NEGLIA, J., REAMAN, G., HOLCENBERG, J. S. & STORK, L. 2002. A randomized comparison of native Escherichia coli asparaginase and polyethylene glycol conjugated asparaginase for treatment of children with newly diagnosed standard-risk acute lymphoblastic leukemia: a Children's Cancer Group study. *Blood,* 99**,** 1986-94.

BARBUL, A., RETTURA, G., LEVENSON, S. M. & SEIFTER, E. 1977. Arginine: a thymotropic and wound-healing promoting agent. *Surg Forum,* 28**,** 101-3.

BARBUL, A., RETTURA, G., LEVENSON, S. M. & SEIFTER, E. 1983. Wound healing and thymotropic effects of arginine: a pituitary mechanism of action. *Am J Clin Nutr,* 37**,** 786-94.

BAYLIN, S. B., HERMAN, J. G., GRAFF, J. R., VERTINO, P. M. & ISSA, J. P. 1998. Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv Cancer Res,* 72**,** 141-96.

BELOUSSOW, K., WANG, L., WU, J., ANN, D. & SHEN, W. C. 2002. Recombinant arginine deiminase as a potential anti-angiogenic agent. *Cancer Lett,* 183**,** 155-62.

BERUTER, J., COLOMBO, J. P. & BACHMANN, C. 1978. Purification and properties of arginase from human liver and erythrocytes. *The Biochemical journal,* 175**,** 449-54.

BILLIAR, T. R. 1995. Nitric oxide. Novel biology with clinical relevance. *Annals of surgery,* 221**,** 339-49.

BOWLES, T. L., KIM, R., GALANTE, J., PARSONS, C. M., VIRUDACHALAM, S., KUNG, H. J. & BOLD, R. J. 2008. Pancreatic cancer cell lines deficient in argininosuccinate synthetase are sensitive to arginine deprivation by arginine deiminase. *Int J Cancer,* 123**,** 1950-5.

BREILLOUT, F., ANTOINE, E. & POUPON, M. F. 1990. Methionine Dependency of Malignant-Tumors - a Possible Approach for Therapy. *Journal of the National Cancer Institute,* 82**,** 1628-1632.

BREILLOUT, F., HADIDA, F., ECHINARDGARIN, P., LASCAUX, V. & POUPON, M. F. 1987. Decreased Rat Rhabdomyosarcoma Pulmonary Metastases in Response to a Low Methionine Diet. *Anticancer Research,* 7**,** 861-867.

BROOME, J. D. 1961. Evidence That L-Asparaginase Activity of Guinea Pig Serum Is Responsible for Its Antilymphoma Effects. *Nature,* 191**,** 1114-&.

BROOME, J. D. 1963. Evidence That L-Asparaginase of Guinea Pig Serum Is Responsible for Its Antilymphoma Effects .1. Properties of L-Asparaginase of Guinea Pig Serum in Relation to Those of Antilymphoma Substance. *Journal of Experimental Medicine,* 118**,** 99-&.

CANTOR, J. R., PANAYIOTOU, V., AGNELLO, G., GEORGIOU, G. & STONE, E. M. 2012a. Engineering reduced-immunogenicity enzymes for amino acid depletion therapy in cancer. *Methods Enzymol,* 502**,** 291-319.

CANTOR, J. R., PANAYIOTOU, V., AGNELLO, G., GEORGIOU, G. & STONES, E. M. 2012b. Engineering Reduced-Immunogenicity Enzymes for Amino Acid Depletion Therapy in Cancer. *Methods in Enzymology, Vol 502: Protein Engineering for Therapeutics, Pt A,* 502**,** 291-319.

CANTOR, J. R., YOO, T. H., DIXIT, A., IVERSON, B. L., FORSTHUBER, T. G. & GEORGIOU, G. 2011. Therapeutic enzyme deimmunization by combinatorial T-cell epitope removal using neutral drift. *Proc Natl Acad Sci U S A,* 108**,** 1272-7.

CASIANOCOLON, A. & MARQUIS, R. E. 1988. Role of the Arginine Deiminase System in Protecting Oral Bacteria and an Enzymatic Basis for Acid Tolerance. *Applied and Environmental Microbiology,* 54**,** 1318-1324.

CASTILLO, L., CHAPMAN, T. E., SANCHEZ, M., YU, Y. M., BURKE, J. F., AJAMI, A. M., VOGT, J. & YOUNG, V. R. 1993. Plasma arginine and citrulline kinetics in adults given adequate and arginine-free diets. *Proc Natl Acad Sci U S A,* 90**,** 7749-53.

CAVALLI, R. C., BURKE, C. J., KAWAMOTO, S., SOPRANO, D. R. & ASH, D. E. 1994. Mutagenesis of rat liver arginase expressed in Escherichia coli: role of conserved histidines. *Biochemistry,* 33**,** 10652-7.

CELLARIER, E., DURANDO, X., VASSON, M. P., FARGES, M. C., DEMIDEN, A., MAURIZIS, J. C., MADELMONT, J. C. & CHOLLET, P. 2003a. Methionine dependency and cancer treatment. *Cancer Treat Rev,* 29**,** 489-99.

CELLARIER, E., DURANDO, X., VASSON, M. P., FARGES, M. C., DEMIDEN, A., MAURIZIS, J. C., MADELMONT, J. C. & CHOLLET, P. 2003b. Methionine dependency and cancer treatment. *Cancer Treatment Reviews,* 29**,** 489-499.

CERQUEIRA, N. M. F. S. A., FERNANDES, P. A. & RAMOS, M. J. 2011. Computational Mechanistic Studies Addressed to the Transimination Reaction Present in All Pyridoxal 5 '-Phosphate-Requiring Enzymes. *Journal of Chemical Theory and Computation,* 7**,** 1356-1368.

CHENG, P. N., LAM, T. L., LAM, W. M., TSUI, S. M., CHENG, A. W., LO, W. H. & LEUNG, Y. C. 2007. Pegylated recombinant human arginase (rhArg-peg5,000mw) inhibits the in vitro and in vivo proliferation of human hepatocellular carcinoma through arginine depletion. *Cancer Res,* 67**,** 309-17.

CHRISTA, L., KERSUAL, J., AUGE, J. & PERIGNON, J. L. 1986. Salvage of 5'-Deoxy-5'-Methylthioadenosine and L-Homocysteine into Methionine in Cells Cultured in a Methionine-Free Medium - a Study of Methionine-Dependence. *Biochemical and Biophysical Research Communications,* 135**,** 131-138.

CHRISTIANSON, D. W. & COX, J. D. 1999. Catalysis by metal-activated hydroxide in zinc and manganese metalloenzymes. *Annu Rev Biochem,* 68**,** 33-57.

COVINI, D., TARDITO, S., BUSSOLATI, O., CHIARELLI, L. R., PASQUETTO, M. V., DIGILIO, R., VALENTINI, G. & SCOTTI, C. 2012. Expanding targets for a metabolic therapy of cancer: L-asparaginase. *Recent Pat Anticancer Drug Discov,* 7**,** 4-13.

COX, J. D., CAMA, E., COLLELUORI, D. M., PETHE, S., BOUCHER, J. L., MANSUY, D., ASH, D. E. & CHRISTIANSON, D. W. 2001. Mechanistic and metabolic inferences from the binding of substrate analogues and products to arginase. *Biochemistry,* 40**,** 2689-701.

CURRIE, G. A., GYURE, L. & CIFUENTES, L. 1979. Microenvironmental arginine depletion by macrophages in vivo. *Br J Cancer,* 39**,** 613-20.

DAS, K., BUTLER, G. H., KWIATKOWSKI, V., CLARK, A. D., JR., YADAV, P. & ARNOLD, E. 2004. Crystal structures of arginine deiminase with covalent reaction intermediates; implications for catalytic mechanism. *Structure,* 12**,** 657-67.

DAVIS, C. D. & UTHUS, E. O. 2004. DNA methylation, cancer susceptibility, and nutrient interactions. *Exp Biol Med (Maywood),* 229**,** 988-95.

DEJONG, P. J. 1972. L-Asparaginase Production by Streptomyces-Griseus. *Applied Microbiology,* 23**,** 1163-&.

DELAGE, B., LUONG, P., MAHARAJ, L., O'RIAIN, C., SYED, N., CROOK, T., HATZIMICHAEL, E., PAPOUDOU-BAI, A., MITCHELL, T. J., WHITTAKER, S. J., CERIO, R., GRIBBEN, J., LEMOINE, N., BOMALASKI, J., LI, C. F., JOEL, S., FITZGIBBON, J., CHEN, L. T. & SZLOSAREK, P. W. 2012. Promoter methylation of argininosuccinate synthetase-1 sensitises lymphomas to arginine deiminase treatment, autophagy and caspase-dependent apoptosis. *Cell Death Dis,* 3**,** e342.

DILLON, B. J., PRIETO, V. G., CURLEY, S. A., ENSOR, C. M., HOLTSBERG, F. W., BOMALASKI, J. S. & CLARK, M. A. 2004. Incidence and distribution of argininosuccinate synthetase deficiency in human cancers: a method for identifying cancers sensitive to arginine deprivation. *Cancer,* 100**,** 826-33.

DIZIKES, G. J., GRODY, W. W., KERN, R. M. & CEDERBAUM, S. D. 1986. Isolation of human liver arginase cDNA and demonstration of nonhomology between the two human arginase genes. *Biochemical and biophysical research communications,* 141**,** 53-9.

DOLOWY, W. C., HENSON, D., CORNET, J. A. & SELLIN, H. 1966. Toxic and Antineoplastic Effects of L-Asparaginase - Study of Mice with Lymphoma and Normal Monkeys and Report on a Child with Leukemia. *Cancer,* 19**,** 1813-&.

ENSOR, C. M., HOLTSBERG, F. W., BOMALASKI, J. S. & CLARK, M. A. 2002. Pegylated arginine deiminase (ADI-SS PEG20,000 mw) inhibits human melanomas and hepatocellular carcinomas in vitro and in vivo. *Cancer Res,* 62**,** 5443-50.

ESAKI, N. & SODA, K. 1987. L-methionine gamma-lyase from Pseudomonas putida and Aeromonas. *Methods Enzymol,* 143**,** 459-65.

FEUN, L. & SAVARAJ, N. 2006. Pegylated arginine deiminase: a novel anticancer enzyme agent. *Expert Opin Investig Drugs,* 15**,** 815-22.

FEUN, L., YOU, M., WU, C. J., KUO, M. T., WANGPAICHITR, M., SPECTOR, S. & SAVARAJ, N. 2008. Arginine deprivation as a targeted therapy for cancer. *Current Pharmaceutical Design,* 14**,** 1049-1057.

FEUN, L. G., KUO, M. T. & SAVARAJ, N. 2015. Arginine deprivation in cancer therapy. *Curr Opin Clin Nutr Metab Care,* 18**,** 78-82.

FEUN, L. G., MARINI, A., WALKER, G., ELGART, G., MOFFAT, F., RODGERS, S. E., WU, C. J., YOU, M., WANGPAICHITR, M., KUO, M. T., SISSON, W., JUNGBLUTH, A. A., BOMALASKI, J. & SAVARAJ, N. 2012. Negative argininosuccinate synthetase expression in melanoma tumours may predict clinical benefit from arginine-depleting therapy with pegylated arginine deiminase. *Br J Cancer,* 106**,** 1481-5.

FISKERSTRAND, T., CHRISTENSEN, B., TYSNES, O. B., UELAND, P. M. & REFSUM, H. 1994. Development and Reversion of Methionine Dependence in a Human Glioma Cell-Line - Relation to Homocysteine Remethylation and Cobalamin Status. *Cancer Research,* 54**,** 4899-4906.

FUKUMOTO, M., KUDOU, D., MURANO, S., SHIBA, T., SATO, D., TAMURA, T., HARADA, S. & INAGAKI, K. 2012. The Role of Amino Acid Residues in the Active Site of L-Methionine γ-lyase from Pseudomonas putida. *Bioscience, Biotechnology, and Biochemistry,* 76**,** 1275-1284.

GALKIN, A., LU, X., DUNAWAY-MARIANO, D. & HERZBERG, O. 2005. Crystal structures representing the Michaelis complex and the thiouronium reaction intermediate of Pseudomonas aeruginosa arginine deiminase. *J Biol Chem,* 280**,** 34080-7.

GEIMAN, T. M. & MUEGGE, K. 2010. DNA methylation in early development. *Mol Reprod Dev,* 77**,** 105-13.

GESTO, D. S., CERQUEIRA, N. M. F. S. A., FERNANDES, P. A. & RAMOS, M. J. 2013. Unraveling the Enigmatic Mechanism of L-Asparaginase II with QM/QM Calculations. *Journal of the American Chemical Society,* 135**,** 7146-7158.

GLAZER, E. S., PICCIRILLO, M., ALBINO, V., DI GIACOMO, R., PALAIA, R., MASTRO, A. A., BENEDUCE, G., CASTELLO, G., DE ROSA, V., PETRILLO, A., ASCIERTO, P. A., CURLEY, S. A. & IZZO, F. 2010. Phase II study of pegylated arginine deiminase for nonresectable and metastatic hepatocellular carcinoma. *J Clin Oncol,* 28**,** 2220-6.

GLAZER, E. S., STONE, E. M., ZHU, C., MASSEY, K. L., HAMIR, A. N. & CURLEY, S. A. 2011. Bioengineered human arginase I with enhanced activity and stability controls hepatocellular and pancreatic carcinoma xenografts. *Translational oncology,* 4**,** 138-46.

GONG, H., ZOLZER, F., VON RECKLINGHAUSEN, G., HAVERS, W. & SCHWEIGERER, L. 2000. Arginine deiminase inhibits proliferation of human leukemia cells more potently than asparaginase by inducing cell cycle arrest and apoptosis. *Leukemia,* 14**,** 826-9.

GOSEKI, N., ONODERA, T., KOIKE, M. & KOSAKI, G. 1987. Inhibitory Effect of L-Methionine-Deprived Amino-Acid Imbalance Using Total Parenteral-Nutrition on Growth of Ascites Hepatoma in Rats. *Tohoku Journal of Experimental Medicine,* 151**,** 191-200.

GUEDES, R. L., PROSDOCIMI, F., FERNANDES, G. R., MOURA, L. K., RIBEIRO, H. A. & ORTEGA, J. M. 2011. Amino acids biosynthesis and nitrogen assimilation pathways: a great genomic deletion during eukaryotes evolution. *BMC Genomics,* 12 Suppl 4**,** S2.

GUO, H., LISHKO, V. K., HERRERA, H., GROCE, A., KUBOTA, T. & HOFFMAN, R. M. 1993a. Therapeutic tumor-specific cell cycle block induced by methionine starvation in vivo. *Cancer Res,* 53**,** 5676-9.

GUO, H. Y., LISHKO, V. K., HERRERA, H., GROCE, A., KUBOTA, T. & HOFFMAN, R. M. 1993b. Therapeutic Tumor-Specific Cell-Cycle Block Induced by Methionine Starvation in-Vivo. *Cancer Research,* 53**,** 5676-5679.

HAINES, R. J., PENDLETON, L. C. & EICHLER, D. C. 2011. Argininosuccinate synthase: at the center of arginine metabolism. *International journal of biochemistry and molecular biology,* 2**,** 8-23.

HENRIKSEN, L. T., HARILA-SAARI, A., RUUD, E., ABRAHAMSSON, J., PRUUNSILD, K., VAITKEVICIENE, G., JONSSON, O. G., SCHMIEGELOW, K., HEYMAN, M., SCHRODER, H., ALBERTSEN, B. K., NORDIC SOCIETY OF PAEDIATRIC, H. & ONCOLOGY, G. 2015. PEG-asparaginase allergy in children with acute lymphoblastic leukemia in the NOPHO ALL2008 protocol. *Pediatr Blood Cancer,* 62**,** 427-33.

HERNANDEZ, C. P., MORROW, K., LOPEZ-BARCONS, L. A., ZABALETA, J., SIERRA, R., VELASCO, C., COLE, J. & RODRIGUEZ, P. C. 2010. Pegylated arginase I: a potential therapeutic approach in T-ALL. *Blood,* 115**,** 5214-21.

HERZFELD, A. & RAPER, S. M. 1976. The heterogeneity of arginases in rat tissues. *The Biochemical journal,* 153**,** 469-78.

HILL, J. M., ROBERTS, J., LOEB, E., KHAN, A., MACLELLAN, A. & HILL, R. W. 1967. L-asparaginase therapy for leukemia and other malignant neoplasms. Remission in human leukemia. *JAMA,* 202**,** 882-8.

HO, E., BEAVER, L. M., WILLIAMS, D. E. & DASHWOOD, R. H. 2011. Dietary factors and epigenetic regulation for prostate cancer prevention. *Adv Nutr,* 2**,** 497-510.

HOFFMAN, R. M. 1984. Altered methionine metabolism, DNA methylation and oncogene expression in carcinogenesis. A review and synthesis. *Biochim Biophys Acta,* 738**,** 49-87.

HOFFMAN, R. M. 1985. Altered Methionine Metabolism and Transmethylation in Cancer. *Anticancer Research,* 5**,** 1-30.

HORI, H., TAKABAYASHI, K., ORVIS, L., CARSON, D. A. & NOBORI, T. 1996. Gene cloning and characterization of Pseudomonas putida L-methionine-alpha-deamino-gamma-mercaptomethane-lyase. *Cancer Research,* 56**,** 2116-2122.

HORN, F. 1933. The breakdown of arginine to citrulline by Bacillus pyocyaneus. *Hoppe Seylers Z Physiol Chem,* 216**,** 244-247.

HOSHIYA, Y., GUO, H., KUBOTA, T., INADA, T., ASANUMA, F., YAMADA, Y., KOH, J., KITAJIMA, M. & HOFFMAN, R. M. 1995. Human tumors are methionine dependent in vivo. *Anticancer Res,* 15**,** 717-8.

HOSHIYA, Y., KUBOTA, T., INADA, T., KITAJIMA, M. & HOFFMAN, R. M. 1997. Methionine-depletion modulates the efficacy of 5-fluorouracil in human gastric cancer in nude mice. *Anticancer Res,* 17**,** 4371-5.

HU, J. & CHEUNG, N.-K. V. 2009. Methionine depletion with recombinant methioninase: In vitro and in vivo efficacy against neuroblastoma and its synergism with chemotherapeutic drugs. *International journal of cancer. Journal international du cancer,* 124**,** 1700-1706.

HUANG, C. C., TSAI, S. T., KUO, C. C., CHANG, J. S., JIN, Y. T., CHANG, J. Y. & HSIAO, J. R. 2012. Arginine deprivation as a new treatment strategy for head and neck cancer. *Oral Oncol,* 48**,** 1227-35.

HUANG, H. Y., WU, W. R., WANG, Y. H., WANG, J. W., FANG, F. M., TSAI, J. W., LI, S. H., HUNG, H. C., YU, S. C., LAN, J., SHIUE, Y. L., HSING, C. H., CHEN, L. T. & LI, C. F. 2013. ASS1 as a novel tumor suppressor gene in myxofibrosarcomas: aberrant loss via epigenetic DNA methylation confers aggressive phenotypes, negative prognostic impact, and therapeutic relevance. *Clin Cancer Res,* 19**,** 2861-72.

ILIES, M., DI COSTANZO, L., DOWLING, D. P., THORN, K. J. & CHRISTIANSON, D. W. 2011. Binding of alpha,alpha-disubstituted amino acids to arginase suggests new avenues for inhibitor design. *J Med Chem,* 54**,** 5432-43.

INOUE, H., INAGAKI, K., ADACHI, N., TAMURA, T., ESAKI, N., SODA, K. & TANAKA, H. 2000. Role of tyrosine 114 of L-methionine gamma-lyase from Pseudomonas putida. *Biosci Biotechnol Biochem,* 64**,** 2336-43.

IZZO, F., MARRA, P., BENEDUCE, G., CASTELLO, G., VALLONE, P., DE ROSA, V., CREMONA, F., ENSOR, C. M., HOLTSBERG, F. W., BOMALASKI, J. S., CLARK, M. A., NG, C. & CURLEY, S. A. 2004. Pegylated arginine deiminase treatment of patients with unresectable hepatocellular carcinoma: results from phase I/II studies. *J Clin Oncol,* 22**,** 1815-22.

JEANBLANC, M., MOUSLI, M., HOPFNER, R., BATHAMI, K., MARTINET, N., ABBADY, A. Q., SIFFERT, J. C., MATHIEU, E., MULLER, C. D. & BRONNER, C. 2005. The retinoblastoma gene and its product are targeted by ICBP90: a key mechanism in the G1/S transition during the cell cycle. *Oncogene,* 24**,** 7337-45.

JENKINSON, C. P., GRODY, W. W. & CEDERBAUM, S. D. 1996. Comparative properties of arginases. *Comparative biochemistry and physiology Part B, Biochemistry & molecular biology,* 114**,** 107-32.

KAFKEWITZ, D. & GOODMAN, D. 1974. L-Asparaginase production by the rumen anaerobe Vibrio succinogenes. *Appl Microbiol,* 27**,** 206-9.

KANYO, Z. F., SCOLNICK, L. R., ASH, D. E. & CHRISTIANSON, D. W. 1996. Structure of a unique binuclear manganese cluster in arginase. *Nature,* 383**,** 554-7.

KEATING, M. J., HOLMES, R., LERNER, S. & HO, D. H. 1993. L-asparaginase and PEG asparaginase--past, present, and future. *Leuk Lymphoma,* 10 Suppl**,** 153-7.

KELLY, M. P., JUNGBLUTH, A. A., WU, B. W., BOMALASKI, J., OLD, L. J. & RITTER, G. 2012. Arginine deiminase PEG20 inhibits growth of small cell lung cancers lacking expression of argininosuccinate synthetase. *Br J Cancer,* 106**,** 324-32.

KENYON, S. H., WATERFIELD, C. J., TIMBRELL, J. A. & NICOLAOU, A. 2002. Methionine synthase activity and sulphur amino acid levels in the rat liver tumour cells HTC and Phi-1. *Biochemical Pharmacology,* 63**,** 381-391.

KHANGULOV, S. V., SOSSONG, T. M., JR., ASH, D. E. & DISMUKES, G. C. 1998. L-arginine binding to liver arginase requires proton transfer to gateway residue His141 and coordination of the guanidinium group to the dimanganese(II,II) center. *Biochemistry,* 37**,** 8539-50.

KIDD, J. G. 1953. Regression of Transplanted Lymphomas Induced Invivo by Means of Normal Guinea Pig Serum .1. Course of Transplanted Cancers of Various Kinds in Mice and Rats Given Guinea Pig Serum, Horse Serum, or Rabbit Serum. *Journal of Experimental Medicine,* 98**,** 565-&.

KIM, A., OH, J. H., PARK, J. M. & CHUNG, A. S. 2007. Methylselenol generated from selenomethionine by methioninase downregulates integrin expression and induces caspase-mediated apoptosis of B16F10 melanoma cells. *J Cell Physiol,* 212**,** 386-400.

KIM, R. H., COATES, J. M., BOWLES, T. L., MCNERNEY, G. P., SUTCLIFFE, J., JUNG, J. U., GANDOUR-EDWARDS, R., CHUANG, F. Y., BOLD, R. J. & KUNG, H. J. 2009. Arginine deiminase as a novel therapy for prostate cancer induces autophagy and caspase-independent apoptosis. *Cancer Res,* 69**,** 700-8.

KLEIN, D. & MORRIS, D. R. 1978. Increased arginase activity during lymphocyte mitogenesis. *Biochemical and biophysical research communications,* 81**,** 199-204.

KOBRINSKY, N. L., SPOSTO, R., SHAH, N. R., ANDERSON, J. R., DELAAT, C., MORSE, M., WARKENTIN, P., GILCHRIST, G. S., COHEN, M. D., SHINA, D. & MEADOWS, A. T. 2001. Outcomes of treatment of children and adolescents with recurrent non-Hodgkin's lymphoma and Hodgkin's disease with dexamethasone, etoposide, cisplatin, cytarabine, and l-asparaginase, maintenance chemotherapy, and transplantation: Children's Cancer Group Study CCG-5912. *J Clin Oncol,* 19**,** 2390-6.

KOKKINAKIS, D. M., HOFFMAN, R. M., FRENKEL, E. P., WICK, J. B., HAN, Q. H., XU, M. X., TAN, Y. Y. & SCHOLD, S. C. 2001. Synergy between methionine stress and chemotherapy in the treatment of brain tumor xenografts in athymic mice. *Cancer Research,* 61**,** 4017-4023.

KREBS, H. 1973. The discovery of the ornithine cycle of urea synthesis. *Biochemical Education,* 1**,** 19-23.

KREIS, W. & HESSION, C. 1973. Isolation and purification of L-methionine-alpha-deamino-gamma-mercaptomethane-lyase (L-methioninase) from Clostridium sporogenes. *Cancer Res,* 33**,** 1862-5.

KUDOU, D., MISAKI, S., YAMASHITA, M., TAMURA, T., ESAKI, N. & INAGAKI, K. 2008. The role of cysteine 116 in the active site of the antitumor enzyme L-methionine gamma-lyase from Pseudomonas putida. *Biosci Biotechnol Biochem,* 72**,** 1722-30.

KUDOU, D., MISAKI, S., YAMASHITA, M., TAMURA, T., TAKAKURA, T., YOSHIOKA, T., YAGI, S., HOFFMAN, R. M., TAKIMOTO, A., ESAKI, N. & INAGAKI, K. 2007a. Structure of the antitumour enzyme L-methionine gamma-lyase from Pseudomonas putida at 1.8 A resolution. *J Biochem,* 141**,** 535-44.

KUDOU, D., MISAKI, S., YAMASHITA, M., TAMURA, T., TAKAKURA, T., YOSHIOKA, T., YAGI, S., HOFFMAN, R. M., TAKIMOTO, A., ESAKI, N. & INAGAKI, K. 2007b. Structure of the antitumour enzyme L-methionine gamma-lyase from pseudomonas putida at 1.8 angstrom resolution. *Journal of Biochemistry,* 141**,** 535-544.

KUZNETSOV, N. A., FALEEV, N. G., KUZNETSOVA, A. A., MOROZOVA, E. A., REVTOVICH, S. V., ANUFRIEVA, N. V., NIKULIN, A. D., FEDOROVA, O. S. & DEMIDKINA, T. V. 2015. Pre-steady-state kinetic and structural analysis of interaction of methionine gamma-lyase from Citrobacter freundii with inhibitors. *J Biol Chem,* 290**,** 671-81.

LAVULO, L. T., SOSSONG, T. M., BRIGHAM-BURKE, M. R., DOYLE, M. L., COX, J. D., CHRISTIANSON, D. W. & ASH, D. E. 2001. Subunit-subunit interactions in trimeric arginase - Generation of active monomers by mutation of a single amino acid. *Journal of Biological Chemistry,* 276**,** 14242-14248.

LEOPOLDINI, M., RUSSO, N. & TOSCANO, M. 2009. Determination of the catalytic pathway of a manganese arginase enzyme through density functional investigation. *Chemistry (Weinheim an der Bergstrasse, Germany),* 15**,** 8026-36.

LIU, J., MA, J., WU, Z., LI, W., ZHANG, D., HAN, L., WANG, F., REINDL, K. M., WU, E. & MA, Q. 2014. Arginine deiminase augments the chemosensitivity of argininosuccinate synthetase-deficient pancreatic cancer cells to gemcitabine via inhibition of NF-kappaB signaling. *BMC Cancer,* 14**,** 686.

LU, X., GALKIN, A., HERZBERG, O. & DUNAWAY-MARIANO, D. 2004. Arginine deiminase uses an active-site cysteine in nucleophilic catalysis of L-arginine hydrolysis. *J Am Chem Soc,* 126**,** 5374-5.

LU, X., LI, L., WU, R., FENG, X., LI, Z., YANG, H., WANG, C., GUO, H., GALKIN, A., HERZBERG, O., MARIANO, P. S., MARTIN, B. M. & DUNAWAY-MARIANO, D. 2006. Kinetic analysis of Pseudomonas aeruginosa arginine deiminase mutants and alternate substrates provides insight into structural determinants of function. *Biochemistry,* 45**,** 1162-72.

MAMAEVA, D. V., MOROZOVA, E. A., NIKULIN, A. D., REVTOVICH, S. V., NIKONOV, S. V., GARBER, M. B. & DEMIDKINA, T. V. 2005. Structure of Citrobacter freundii L-methionine gamma-lyase. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 61**,** 546-9.

MCKIE, A. E., EDLIND, T., WALKER, J., MOTTRAM, J. C. & COOMBS, G. H. 1998. The primitive protozoon Trichomonas vaginalis contains two methionine gamma-lyase genes that encode members of the gamma-family of pyridoxal 5'-phosphate-dependent enzymes. *J Biol Chem,* 273**,** 5549-56.

MIKI, K., XU, M., AN, Z., WANG, X., YANG, M., AL-REFAIE, W., SUN, X., BARANOV, E., TAN, Y., CHISHIMA, T., SHIMADA, H., MOOSSA, A. R. & HOFFMAN, R. M. 2000a. Survival efficacy of the combination of the methioninase gene and methioninase in a lung cancer orthotopic model. *Cancer Gene Ther,* 7**,** 332-8.

MIKI, K., XU, M., GUPTA, A., BA, Y., TAN, Y., AL-REFAIE, W., BOUVET, M., MAKUUCHI, M., MOOSSA, A. R. & HOFFMAN, R. M. 2001. Methioninase cancer gene therapy with selenomethionine as suicide prodrug substrate. *Cancer Res,* 61**,** 6805-10.

MIKI, K., XU, M. X., AN, Z. L., WANG, X. E., YANG, M., AL-REFAIE, W., SUN, X. H., BARANOV, E., TAN, Y. Y., CHISHIMA, T., SHIMADA, H., MOOSSA, A. R. & HOFFMAN, R. M. 2000b. Survival efficacy of the combination of the methioninase gene and methioninase in a lung cancer orthotopic model. *Cancer Gene Therapy,* 7**,** 332-338.

MILLIS, R. M., DIYA, C. A., REYNOLDS, M. E., DEHKORDI, O. & BOND, V. 1998. Growth inhibition of subcutaneously transplanted hepatomas without cachexia by alteration of the dietary arginine-methionine balance. *Nutrition and Cancer-an International Journal,* 31**,** 49-55.

MORRIS, S. M., JR., BHAMIDIPATI, D. & KEPKA-LENHART, D. 1997. Human type II arginase: sequence analysis and tissue-specific expression. *Gene,* 193**,** 157-61.

MOTOSHIMA, H., INAGAKI, K., KUMASAKA, T., FURUICHI, M., INOUE, H., TAMURA, T., ESAKI, N., SODA, K., TANAKA, N., YAMAMOTO, M. & TANAKA, H. 2000. Crystal structure of the pyridoxal 5'-phosphate dependent L-methionine gamma-lyase from Pseudomonas putida. *J Biochem,* 128**,** 349-54.

MULLER, H. J. & BOOS, J. 1998. Use of L-asparaginase in childhood ALL. *Critical reviews in oncology/hematology,* 28**,** 97-113.

NAKAYAMA, T., ESAKI, N., SUGIE, K., BERESOV, T. T., TANAKA, H. & SODA, K. 1984. Purification of bacterial L-methionine gamma-lyase. *Anal Biochem,* 138**,** 421-4.

NI, Y., SCHWANEBERG, U. & SUN, Z. H. 2008. Arginine deiminase, a potential anti-tumor drug. *Cancer Lett,* 261**,** 1-11.

NOBORI, T., SZINAI, I., AMOX, D., PARKER, B., OLOPADE, O. I., BUCHHAGEN, D. L. & CARSON, D. A. 1993. Methylthioadenosine phosphorylase deficiency in human non-small cell lung cancers. *Cancer Res,* 53**,** 1098-101.

NOZAKI, T., ALI, V. & TOKORO, M. 2005. Sulfur-containing amino acid metabolism in parasitic protozoa. *Adv Parasitol,* 60**,** 1-99.

OLIVEIRA, E. F., CERQUEIRA, N. M. F. S. A., FERNANDES, P. A. & RAMOS, M. J. 2011. Mechanism of Formation of the Internal Aldimine in Pyridoxal 5 '-Phosphate-Dependent Enzymes. *Journal of the American Chemical Society,* 133**,** 15496-15505.

OTT, P. A., CARVAJAL, R. D., PANDIT-TASKAR, N., JUNGBLUTH, A. A., HOFFMAN, E. W., WU, B. W., BOMALASKI, J. S., VENHAUS, R., PAN, L., OLD, L. J., PAVLICK, A. C. & WOLCHOK, J. D. 2013. Phase I/II study of pegylated arginine deiminase (ADI-PEG 20) in patients with advanced melanoma. *Invest New Drugs,* 31**,** 425-34.

PARK, I. S., KANG, S. W., SHIN, Y. J., CHAE, K. Y., PARK, M. O., KIM, M. Y., WHEATLEY, D. N. & MIN, B. H. 2003. Arginine deiminase: a potential inhibitor of angiogenesis and tumour growth. *British Journal of Cancer,* 89**,** 907-914.

PETERSON, R. E. & CIEGLER, A. 1969. L-Asparaginase Production by Various Bacteria. *Applied Microbiology,* 17**,** 929-&.

POIRSON-BICHAT, F., GONCALVES, R. A. B., MICCOLI, L., DUTRILLAUX, B. & POUPON, M. F. 2000. Methionine depletion enhances the antitumoral efficacy of cytotoxic agents in drug-resistant human tumor xenografts. *Clinical Cancer Research,* 6**,** 643-653.

QIU, F., CHEN, Y. R., LIU, X., CHU, C. Y., SHEN, L. J., XU, J., GAUR, S., FORMAN, H. J., ZHANG, H., ZHENG, S., YEN, Y., HUANG, J., KUNG, H. J. & ANN, D. K. 2014. Arginine starvation impairs mitochondrial respiratory function in ASS1-deficient breast cancer cells. *Sci Signal,* 7**,** ra31.

QIU, F., HUANG, J. & SUI, M. 2015. Targeting arginine metabolism pathway to treat arginine-dependent cancers. *Cancer Lett,* 364**,** 1-7.

RAMANI, K., YANG, H., KUHLENKAMP, J., TOMASI, L., TSUKAMOTO, H., MATO, J. M. & LU, S. C. 2010. Changes in the expression of methionine adenosyltransferase genes and S-adenosylmethionine homeostasis during hepatic stellate cell activation. *Hepatology,* 51**,** 986-95.

RECZKOWSKI, R. S. & ASH, D. E. 1992. EPR evidence for binuclear manganese (II) centers in rat liver arginase. *Journal of the American Chemical Society,* 114**,** 10992-10994.

REEDS, P. J. 2000. Dispensable and indispensable amino acids for humans. *J Nutr,* 130**,** 1835S-40S.

RICHARDS, N. G. J. & KILBERG, M. S. 2006. Asparagine synthetase chemotherapy. *Annual Review of Biochemistry,* 75**,** 629-654.

ROGERS, Q. R. & VISEK, W. J. 1985. Metabolic role of urea cycle intermediates: nutritional and clinical aspects. Introduction. *The Journal of nutrition,* 115**,** 505-8.

RYTTING, M. 2010a. Peg-asparaginase for acute lymphoblastic leukemia. *Expert Opin Biol Ther,* 10**,** 833-9.

RYTTING, M. 2010b. Peg-asparaginase for acute lymphoblastic leukemia. *Expert Opinion on Biological Therapy,* 10**,** 833-839.

SARQUIS, M. I., OLIVEIRA, E. M., SANTOS, A. S. & COSTA, G. L. 2004. Production of L-asparaginase by filamentous fungi. *Mem Inst Oswaldo Cruz,* 99**,** 489-92.

SATO, D. & NOZAKI, T. 2009. Methionine gamma-lyase: the unique reaction mechanism, physiological roles, and therapeutic applications against infectious diseases and cancers. *IUBMB Life,* 61**,** 1019-28.

SATO, D., YAMAGATA, W., HARADA, S. & NOZAKI, T. 2008. Kinetic characterization of methionine gamma-lyases from the enteric protozoan parasite Entamoeba histolytica against physiological substrates and trifluoromethionine, a promising lead compound against amoebiasis. *FEBS J,* 275**,** 548-60.

SAVARAJ, N., YOU, M., WU, C., WANGPAICHITR, M., KUO, M. T. & FEUN, L. G. 2010. Arginine deprivation, autophagy, apoptosis (AAA) for the treatment of melanoma. *Curr Mol Med,* 10**,** 405-12.

SAVOCA, K. V., ABUCHOWSKI, A., VAN ES, T., DAVIS, F. F. & PALCZUK, N. C. 1979. Preparation of a non-immunogenic arginase by the covalent attachment of polyethylene glycol. *Biochimica et biophysica acta,* 578**,** 47-53.

SCHELLEKENS, H. 2002. Immunogenicity of therapeutic proteins: Clinical implications and future prospects. *Clinical Therapeutics,* 24**,** 1720-1740.

SCOLNICK, L. R., KANYO, Z. F., CAVALLI, R. C., ASH, D. E. & CHRISTIANSON, D. W. 1997. Altering the binuclear manganese cluster of arginase diminishes thermostability and catalytic function. *Biochemistry,* 36**,** 10558-65.

SHARMA, B., SINGH, S. & KANWAR, S. S. 2014. L-methionase: a therapeutic enzyme to treat malignancies. *Biomed Res Int,* 2014**,** 506287.

SHEN, L. J., LIN, W. C., BELOUSSOW, K. & SHEN, W. C. 2003. Resistance to the anti-proliferative activity of recombinant arginine deiminase in cell culture correlates with the endogenous enzyme, argininosuccinate synthetase. *Cancer Letters,* 191**,** 165-170.

SHEN, L. J. & SHEN, W. C. 2006. Drug evaluation: ADI-PEG-20 - a PEGylated arginine deiminase for arginine-auxotrophic cancers. *Current Opinion in Molecular Therapeutics,* 8**,** 240-248.

SIMON-REUSS, I. 1953. Arginase, an antimitotic agent in tissue culture. *Biochimica et biophysica acta,* 11**,** 396-402.

SMIRAGLIA, D. J., RUSH, L. J., FRUHWALD, M. C., DAI, Z., HELD, W. A., COSTELLO, J. F., LANG, J. C., ENG, C., LI, B., WRIGHT, F. A., CALIGIURI, M. A. & PLASS, C. 2001. Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. *Hum Mol Genet,* 10**,** 1413-9.

SMITH, D. W., GANAWAY, R. L. & FAHRNEY, D. E. 1978. Arginine deiminase from Mycoplasma arthritidis. Structure-activity relationships among substrates and competitive inhibitors. *J Biol Chem,* 253**,** 6016-20.

SPECTOR, E. B., RICE, S. C. & CEDERBAUM, S. D. 1983. Immunologic studies of arginase in tissues of normal human adult and arginase-deficient patients. *Pediatric research,* 17**,** 941-4.

STONE, E. M., GLAZER, E. S., CHANTRANUPONG, L., CHERUKURI, P., BREECE, R. M., TIERNEY, D. L., CURLEY, S. A., IVERSON, B. L. & GEORGIOU, G. 2010. Replacing Mn(2+) with Co(2+) in human arginase i enhances cytotoxicity toward l-arginine auxotrophic cancer cell lines. *ACS chemical biology,* 5**,** 333-42.

STORY, M. D., VOEHRINGER, D. W., STEPHENS, L. C. & MEYN, R. E. 1993. L-asparaginase kills lymphoma cells by apoptosis. *Cancer Chemother Pharmacol,* 32**,** 129-33.

SUBHI, A. L., DIEGELMAN, P., PORTER, C. W., TANG, B., LU, Z. J., MARKHAM, G. D. & KRUGER, W. D. 2003. Methylthioadenosine phosphorylase regulates ornithine decarboxylase by production of downstream metabolites. *J Biol Chem,* 278**,** 49868-73.

SUGIMURA, K., OHNO, T., KIMURA, Y., KIMURA, T. & AZUMA, I. 1992. Arginine deiminase gene of an AIDS-associated mycoplasma, Mycoplasma incognitus. *Microbiol Immunol,* 36**,** 667-70.

SUN, X., YANG, Z., LI, S., TAN, Y., ZHANG, N., WANG, X., YAGI, S., YOSHIOKA, T., TAKIMOTO, A., MITSUSHIMA, K., SUGINAKA, A., FRENKEL, E. P. & HOFFMAN, R. M. 2003a. In vivo efficacy of recombinant methioninase is enhanced by the combination of polyethylene glycol conjugation and pyridoxal 5 '-phosphate supplementation. *Cancer Research,* 63**,** 8377-8383.

SUN, X., YANG, Z., LI, S., TAN, Y., ZHANG, N., WANG, X., YAGI, S., YOSHIOKA, T., TAKIMOTO, A., MITSUSHIMA, K., SUGINAKA, A., FRENKEL, E. P. & HOFFMAN, R. M. 2003b. In vivo efficacy of recombinant methioninase is enhanced by the combination of polyethylene glycol conjugation and pyridoxal 5'-phosphate supplementation. *Cancer Res,* 63**,** 8377-83.

SWAIN, A. L., JASKOLSKI, M., HOUSSET, D., RAO, J. K. M. & WLODAWER, A. 1993. Crystal-Structure of Escherichia-Coli L-Asparaginase, an Enzyme Used in Cancer-Therapy. *Proceedings of the National Academy of Sciences of the United States of America,* 90**,** 1474-1478.

SYED, N., LANGER, J., JANCZAR, K., SINGH, P., LO NIGRO, C., LATTANZIO, L., COLEY, H. M., HATZIMICHAEL, E., BOMALASKI, J., SZLOSAREK, P., AWAD, M., O'NEIL, K., RONCAROLI, F. & CROOK, T. 2013. Epigenetic status of argininosuccinate synthetase and argininosuccinate lyase modulates autophagy and cell death in glioblastoma. *Cell Death Dis,* 4**,** e458.

TAKAKU, H., MATSUMOTO, M., MISAWA, S. & MIYAZAKI, K. 1995. Antitumor-Activity of Arginine Deiminase from Mycoplasma-Arginini and Its Growth-Inhibitory Mechanism. *Japanese Journal of Cancer Research,* 86**,** 840-846.

TAKAKU, H., TAKASE, M., ABE, S., HAYASHI, H. & MIYAZAKI, K. 1992. Invivo Antitumor-Activity of Arginine Deiminase Purified from Mycoplasma-Arginini. *International Journal of Cancer,* 51**,** 244-249.

TAN, Y., SUN, X., XU, M., AN, Z., TAN, X., HAN, Q., MILJKOVIC, D. A., YANG, M. & HOFFMAN, R. M. 1998. Polyethylene glycol conjugation of recombinant methioninase for cancer therapy. *Protein Expr Purif,* 12**,** 45-52.

TAN, Y., XU, M., GUO, H., SUN, X., KUBOTA, T. & HOFFMAN, R. M. 1996. Anticancer efficacy of methioninase in vivo. *Anticancer Res,* 16**,** 3931-6.

TAN, Y., XU, M., TAN, X., TAN, X., WANG, X., SAIKAWA, Y., NAGAHAMA, T., SUN, X., LENZ, M. & HOFFMAN, R. M. 1997. Overexpression and large-scale production of recombinant L-methionine-alpha-deamino-gamma-mercaptomethane-lyase for novel anticancer therapy. *Protein Expr Purif,* 9**,** 233-45.

TAN, Y. Y., SUN, X. H., XU, M. X., TAN, X. Z., SASSON, A., RASHIDI, B., HAN, Q. H., TAN, X. Y., WANG, X. E., AN, Z. L., SUN, F. X. & HOFFMAN, R. M. 1999. Efficacy of recombinant methioninase in combination with cisplatin on human colon tumors in nude mice. *Clinical Cancer Research,* 5**,** 2157-2163.

TAN, Y. Y., XU, M. X. & HOFFMAN, R. M. 2010. Broad Selective Efficacy of Recombinant Methioninase and Polyethylene Glycol-modified Recombinant Methioninase on Cancer Cells In Vitro. *Anticancer Research,* 30**,** 1041-1046.

THIVAT, E., DURANDO, X., DEMIDEM, A., FARGES, M. C., RAPP, M., CELLARIER, E., GUENIN, S., D'INCAN, M., VASSON, M. P. & CHOLLET, P. 2007. A methionine-free diet associated with nitrosourea treatment down-regulates methylguanine-DNA methyl transferase activity in patients with metastatic cancer. *Anticancer Res,* 27**,** 2779-83.

TONEY, M. D. 2005. Reaction specificity in pyridoxal phosphate enzymes. *Arch Biochem Biophys,* 433**,** 279-87.

TOSA, T., SANO, R., YAMAMOTO, K., NAKAMURA, M., ANDO, K. & CHIBATA, I. 1971. L-Asparaginase from Proteus-Vulgaris. *Applied Microbiology,* 22**,** 387-&.

TRAWEEK, S. T., RISCOE, M. K., FERRO, A. J., BRAZIEL, R. M., MAGENIS, R. E. & FITCHEN, J. H. 1988. Methylthioadenosine phosphorylase deficiency in acute leukemia: pathologic, cytogenetic, and clinical features. *Blood,* 71**,** 1568-73.

VOCKLEY, J. G., JENKINSON, C. P., SHUKLA, H., KERN, R. M., GRODY, W. W. & CEDERBAUM, S. D. 1996. Cloning and characterization of the human type II arginase gene. *Genomics,* 38**,** 118-23.

VROOMAN, L. M., SUPKO, J. G., NEUBERG, D. S., ASSELIN, B. L., ATHALE, U. H., CLAVELL, L., KELLY, K. M., LAVERDIERE, C., MICHON, B., SCHORIN, M., COHEN, H. J., SALLAN, S. E. & SILVERMAN, L. B. 2010. Erwinia Asparaginase After Allergy to E. coli Asparaginase in Children With Acute Lymphoblastic Leukemia. *Pediatric Blood & Cancer,* 54**,** 199-205.

WHEATLEY, D. N. & CAMPBELL, E. 2003. Arginine deprivation, growth inhibition and tumour cell death: 3. Deficient utilisation of citrulline by malignant cells. *British Journal of Cancer,* 89**,** 573-576.

WISWELL, O. B. 1951. Effects of intraperitoneally injected arginase on growth of mammary carcinoma implants in the mouse. *Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine (New York, N Y ),* 76**,** 588-9.

WOO, M. H., HAK, L. J., STORM, M. C., EVANS, W. E., SANDLUND, J. T., RIVERA, G. K., WANG, B., PUI, C. H. & RELLING, M. V. 1998. Anti-asparaginase antibodies following E. coli asparaginase therapy in pediatric acute lymphoblastic leukemia. *Leukemia,* 12**,** 1527-33.

YAMAMOTO, N., GUPTA, A., XU, M. X., MIKI, K., TSUJIMOTO, Y., TSUCHIYA, H., TOMITA, K., MOOSSA, A. R. & HOFFMAN, R. M. 2003. Methioninase gene therapy with selenomethionine induces apoptosis in bcl-2-overproducing lung cancer cells. *Cancer Gene Therapy,* 10**,** 445-450.

YANG, T. S., LU, S. N., CHAO, Y., SHEEN, I. S., LIN, C. C., WANG, T. E., CHEN, S. C., WANG, J. H., LIAO, L. Y., THOMSON, J. A., WANG-PENG, J., CHEN, P. J. & CHEN, L. T. 2010. A randomised phase II study of pegylated arginine deiminase (ADI-PEG 20) in Asian advanced hepatocellular carcinoma patients. *Br J Cancer,* 103**,** 954-60.

YANG, Z., WANG, J., YOSHIOKA, T., LI, B., LU, Q., LI, S., SUN, X., TAN, Y., YAGI, S., FRENKEL, E. P. & HOFFMAN, R. M. 2004a. Pharmacokinetics, methionine depletion, and antigenicity of recombinant methioninase in primates. *Clin Cancer Res,* 10**,** 2131-8.

YANG, Z. J., SUN, X. H., LI, S. K., TAN, Y. Y., WANG, X., ZHANG, N., YAGI, S., TAKAKURA, T., KOBAYASHI, Y., TAKIMOTO, A., YOSHIOKA, T., SUGINAKA, A., FRENKEL, E. P. & HOFFMAN, R. M. 2004b. Circulating half-life of PEGylated recombinant methioninase holoenzyme is highly dose dependent on cofactor pyridoxal-5 '-phosphate. *Cancer Research,* 64**,** 5775-5778.

YANG, Z. J., WANG, J. H., LU, Q., XU, J. B., KOBAYASHI, Y., TAKAKURA, T., TAKIMOTO, A., YOSHIOKA, T., LIAN, C. G., CHEN, C. M., ZHANG, D. D., ZHANG, Y., LI, S. K., SUN, X. H., TAN, Y. Y., YAGI, S., FRENKEL, E. P. & HOFFMAN, R. M. 2004c. PEGylation confers greatly extended half-life and attenuated immunogenicity to recombinant methioninase in primates. *Cancer Research,* 64**,** 6673-6678.

YOON, C. Y., SHIM, Y. J., KIM, E. H., LEE, J. H., WON, N. H., KIM, J. H., PARK, I. S., YOON, D. K. & MIN, B. H. 2007. Renal cell carcinoma does not express argininosuccinate synthetase and is highly sensitive to arginine deprivation via arginine deiminase. *International Journal of Cancer,* 120**,** 897-905.

YOSHIOKA, T., WADA, T., UCHIDA, N., MAKI, H., YOSHIDA, H., IDE, N., KASAI, H., HOJO, K., SHONO, K., MAEKAWA, R., YAGI, S., HOFFMAN, R. M. & SUGITA, K. 1998. Anticancer efficacy in vivo and in vitro, synergy with 5-fluorouracil, and safety of recombinant methioninase. *Cancer Res,* 58**,** 2583-7.

ZHANG, L., LIU, M., JAMIL, S., HAN, R., XU, G. & NI, Y. 2015. PEGylation and pharmacological characterization of a potential anti-tumor drug, an engineered arginine deiminase originated from Pseudomonas plecoglossicida. *Cancer Lett,* 357**,** 346-54.

ZINGG, J. M. & JONES, P. A. 1997. Genetic and epigenetic aspects of DNA methylation on genome expression, evolution, mutation and carcinogenesis. *Carcinogenesis,* 18**,** 869-882.

# 3. Conference presentations

**Oral Communication:**

Fernandes, H. S., Cerqueira, N. M. F. S. A., *"Computational studies addressed to the catalytic mechanism of Histidine Decarboxylase"*, IJUP'16 – 9º Encontro de Jovens Investigadores da Universidade do Porto, Fundação da Juventude – Palácio das Artes, 2016 February 17th, Porto

**Poster Presentantion:**

Henrique S. Fernandes, Nuno M. F. S. A. Cerqueira, *"Computational studies addressed to the catalytic mechanism of Histidine Decarboxylase"* | Encontro De Jovens Investigadores Em Biologia Computacional Estrutural (EJIBCE) – Instituto Pedro Nunes, Coimbra – 2015, 18th December

Fernandes, Henrique S., Ramos, Maria J., Cerqueira, Nuno M. F. S. A., *"Computational studies addressed to the catalytic mechanism of Histidine Decarboxylase"*, 5º Encontro Português de Jovens Químicos (PYCheM) and 1st European Young Chemists Meeting – Guimarães, 2016 April 23th

# Computational studies addressed to the catalytic mechanism of Histidine Decarboxylase

Henrique S. Fernandes[1‡], Nuno M. F. S. A. Cerqueira[1]

[1] REQUIMTE/UCIBIO, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal
[‡] henriquefer11@gmail.com

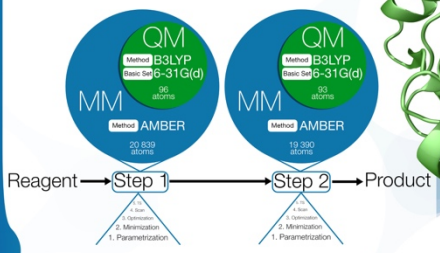
## 1. INTRODUCTION

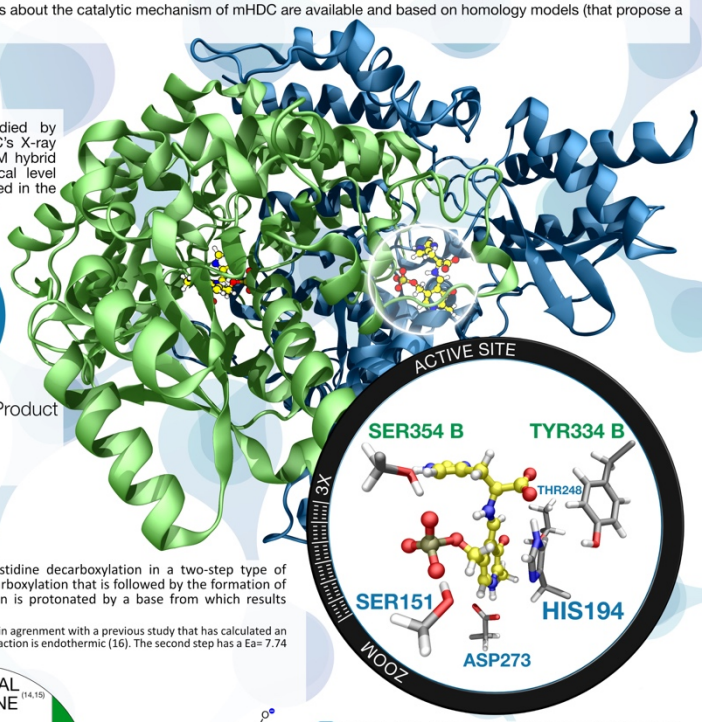
Mammalian histidine decarboxylase (mHDC) is an enzyme that requires pyridoxal-5'-phosphate (PLP) as a cofactor (1). mHDC belongs to group II of PLP-dependent decarboxylases (together with *L*-DOPA and glutamate decarboxylases) and catalyses the decarboxylation of *L*-histidine to histamine (2). Histamine plays a key role in several biological events such as immune response (3), gastric system modulation (4,5) and as a neurotransmitter in the nervous system (6). Several inhibitors for histamine action have been studied in order to treat some diseases such as atopic dermatitis (7), allergies (8), and cancer (9-12).

mHDC has been studied for a long time, but only in 2012 Komori's (13) group was able to determine the X-ray structure of the enzyme and revealed the active site environment. Until date, only hypothesis about the catalytic mechanism of mHDC are available and based on homology models (that propose a different active site configuration).

## 2. METHODOLOGY

The catalytic mechanism of mHDC was studied by computational approaches using a recent mHDC's X-ray structure. In all geometry optimizations a QM/MM hybrid methodology was employed using the theoretical level DFT/6-31G(d)::Amber. More details can be obtained in the below scheme.




## 3. RESULTS

The results have shown that mHDC catalysis the histidine decarboxylation in a two-step type of mechanism. The **first step** involves a histidine (R) decarboxylation that is followed by the formation of a carbanion. Then, in the **second step**, the carbanion is protonated by a base from which results histamine (P).

Our early results indicate that the first step is the limiting one. This is in agrement with a previous study that has calculated an activation energy of 22.6 kcal/mol for this step and showed that the reaction is endothermic (16). The second step has a Ea= 7.74 kcal/mol and it is exothermic by 26.8 kcal/mol.

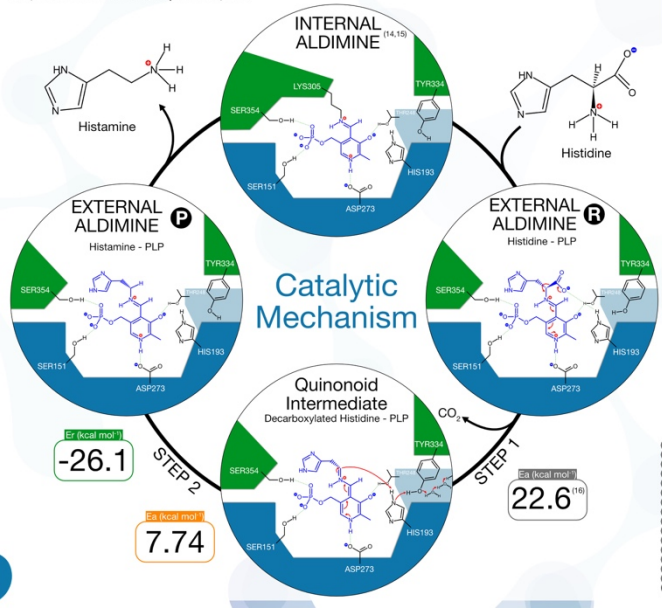



FIG. 1 Structure of the *Homo Sapiens L*-Histidine Decarboxylase dimer (NewCartoon representation; blue subunit A; green subunit B). Natural substrate (*L*-Histidine) linked to the PLP co-factor (CPK representation). Highlight of the active site; Histidine linked to PLP (CPK); active site residues that should be involved in catalysis (Licorice representation). (PDB code: 4E1O adapted) (13)

## 4. FUTURE

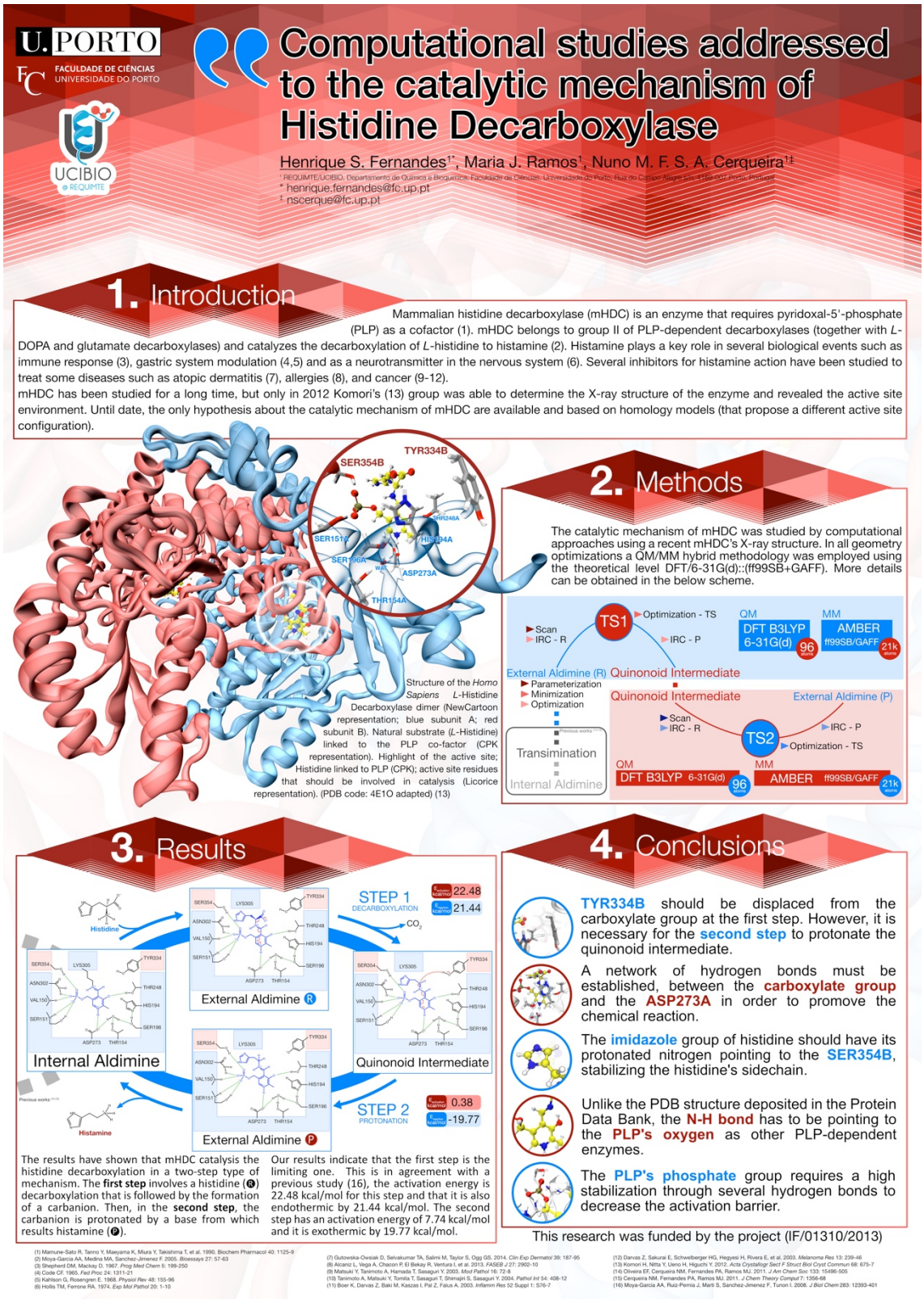
Determination of the activation and reaction energies for the first step.

Improve the level of theory of the final energies through single-point energy calculations.

Computational Alanine Scanning Mutagenesis (CompASM) to search for relevant protein-protein interactions between the two subunits of the mHDC.

(1) Mamune-Sato R, Tanno Y, Maeyama K, Miura Y, Takishima T, et al. 1990. Biochem Pharmacol 40: 1125-9
(2) Moya-Garcia AA, Medina MA, Sanchez-Jimenez F. 2005. *Bioessays* 27: 57-63
(3) Shepherd DM, Mackay D. 1967. *Prog Med Chem* 5: 199-250
(4) Code CF. 1965. *Fed Proc* 24: 1311-21
(5) Kahlson G, Rosengren E. 1968. *Physiol Rev* 48: 155-96
(6) Hollis TM, Fermone RA. 1974. *Exp Mol Pathol* 20: 1-10
(7) Gutowska-Owsiak D, Selvakumar TA, Salimi M, Taylor S, Ogg GS. 2014. *Clin Exp Dermatol* 39: 187-95
(8) Alcaniz L, Vega A, Chacon P, El Bekay R, Ventura I, et al. 2013. *FASEB J* 27: 2902-10
(9) Matsuki Y, Tanimoto A, Hamada T, Sasaguri Y. 2003. *Mod Pathol* 16: 72-8
(10) Tanimoto A, Matsuki Y, Tomita T, Sasaguri T, Shimajiri S, Sasaguri Y. 2004. *Pathol Int* 54: 408-12
(11) Boer K, Darvas Z, Baki M, Kaszas I, Pal Z, Falus A. 2003. *Inflamm Res* 52 Suppl 1: S76-7
(12) Darvas Z, Sakurai E, Schweiberger HG, Hegyesi H, Rivera E, et al. 2003. *Melanoma Res* 13: 239-46
(13) Komori H, Nitta Y, Ueno H, Higuchi Y. 2012. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 68: 675-7
(14) Oliveira EF, Cerqueira NM, Fernandes PA, Ramos MJ. 2011. *J Am Chem Soc* 133: 15496-505
(15) Cerqueira NM, Fernandes PA, Ramos MJ. 2011. *J Chem Theory Comput* 7: 1356-68
(16) Moya-Garcia AA, Ruiz-Pernia JJ, Marti S, Sanchez-Jimenez F, Tunon I. 2008. *J Biol Chem* 283: 12393-401

**Supplementary Fig. 5 – Poster presented at the "Encontro De Jovens Investigadores Em Biologia Computacional Estrutural (EJIBCE)" – Instituto Pedro Nunes, Coimbra – 2015, 18th December**




**Supplementary Fig. 6 – Poster presented at "5º Encontro Português de Jovens Químicos (PYCheM) and 1st European Young Chemists Meeting" – Guimarães, 2016 April 23th**

## 4. Non-academic work

During this project, it was also developed an educational software called “Protein Wars”. This educational game was available during the “14ª Mostra da Universidade do Porto” from 17/03/2016 to 20/03/2016, and its main aim was to show the importance of chemistry and drug development to society in a ludic manner.

“Protein Wars” was developed using the TCL language and it works as a VMD plugin. It was also built an arcade-like machine and configured the buttons in order to make it ready to work with “Protein Wars” and VMD.

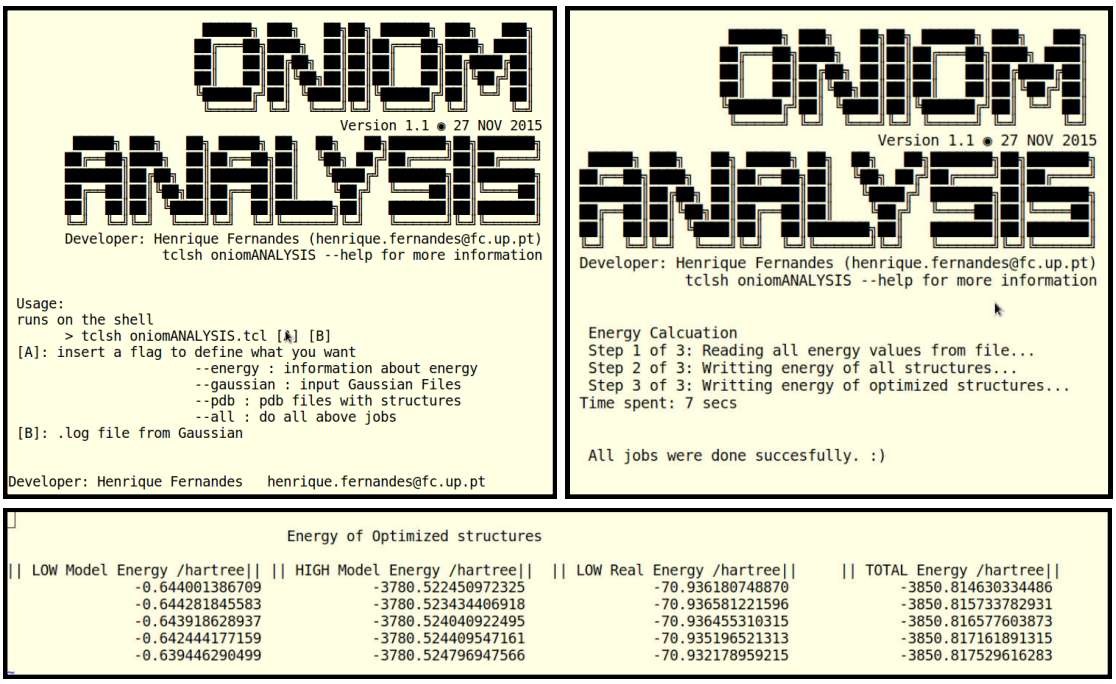
The game has several levels, and in each level, a protein is presented with a short explanation about it and how it is related with a specific disease. The main goal of each level is to target relevant regions of those proteins, which culminate into their destruction. The main objective was to show that several drugs that are commercially available aim the destruction of these enzymes, and this often involves the inactivation of their active sites.

The project will continue to be developed and some improvements and new levels will be implemented in the future.


**Supplementary Fig. 7 – Screenshots of the "Protein Wars" game and photography of the arcade machine construction.**


**Supplementary Fig. 8 – "Protein Wars" game was available during the "14ª Mostra da Universidade do Porto 2016".**

Moreover, two additional scripts were developed. “ONIOM Analysis” (github.com /henriquefer/oniomANALYSIS) was developed to allow an easy manner to extract the structures from Gaussian output files to PDB or Gaussian input files. The script was developed using the TCL language, and it runs in the bash, requiring only the resulting .log file from QM/MM calculations using Gaussian09. Additionally, the script is able to extract the energies of each layer of the QM/MM model, considering all the structures or solely the optimized ones.




**Supplementary Fig. 9 – Screenshots of the “ONIOM Analysis” script.**

A second script, called “ONIOM FROZEN” (github.com/henriquefer/oniom FROZEN) was also developed to handle with the freezing state of the atoms in any Gaussian input file. The script allows the inversion of the freezing status of each atom, unfreezing all atoms of the model, freeze all atoms of the model or freezing the atoms of the residues that were selected by the user (and need to be provided in second file).


**Supplementary Fig. 10 – Screenshot of the "ONIOM Frozen" script.**